Economics Education and Research Consortium

Working Paper Series

ISSN 1561-2422

No 05/12

Socioeconomic determinants of smoking in contemporary Russia

Sergey Arzhenovsky

This project (No. 03-138) was supported
by the Economics Education and Research Consortium

Research area: **Labor Markets and Social Policy**

JEL Classification: C41, D12, I12

ARZHENOVSKY S.V. Socioeconomic determinants of smoking in contemporary Russia. — Moscow: EERC, 2005.

Factors impacting the initiation and termination of smoking, using Cox's proportional hazard model as the econometric tool, on the basis of RLMS and Goskomstat data on tobacco prices, are being investigated in this paper. The model to explain the amount of consumed cigarettes and the composite model with dependence between quitting and tobacco consumption are constructed. It is shown that the price for cigarettes remains the key factor for the beginning and quitting of smoking; the asymmetric influence of price by sorts of cigarettes is being revealed. The addictive character of cigarettes consumption has been confirmed. Gender and age peculiarities of tobacco products consumption have been revealed, especially for teenagers. Opportunities to reduce smoking have been found out through propagation of a healthy way of life.

Keywords. Russia, duration analysis, smoking, hazard model, consumption habits, amount of cigarettes smoked.

Acknowledgements. The author thanks Michael Beenstock, Randall Filer, Irina Denisova for the professional recommendations to realize the project and Anna Lukyanova, Yury Andrienko and participants of EERC workshops 2003–2004 in Kiev and Moscow for useful comments, Valery Smirnov for the help in editing of English text.

Sergey Arzhenovsky
Rostov State Economic University
Chair of mathematical statistics, econometrics and actuarial calculations
B. Sadovaja st. 69, Rostov-on-Don, 344002
Tel.: (863) 261 38 65
Fax: (863) 240 57 29
E-mail: asv1@nm.ru

CONTENTS

NON-TECHNICAL SUMMARY

The present paper investigates the economic aspects of tobacco consumption applying Russian data for the first time. The aim of the research is to analyse the determinants impacting the overcoming of a double-hurdle, *viz.* the beginning and quitting of smoking in contemporary Russia, and also the determinants of tobacco consumption. The importance of the research is determined by the fact that understanding of reasons for smoking is important for develop effective state policy on regulation of tobacco production market.

Models of addiction habit using addictive nature of smoking became the theoretical basis of the research. Kaplan–Meyer's product-limit estimators, Cox's proportional hazard models and panel data regresses were the econometric tools. The data of six waves of Russian Longitudinal Monitoring Survey of economy and health (RLMS) for 1994–2001 and Goskomstat data on prices for cigarettes in regions were applied.

Social and economic changes have led to a wide spreading of a smoking habit at the beginning of the current century. This paper analyses the reasons for beginning and quitting of smoking. Asymmetric influence of cigarette prices by sorts of cigarettes has been revealed. The social problem is being exposed, *i.e.* declining of living standards, when real prices on cheap cigarettes grow, leads to wide practice of a harmful habit of smoking among teenagers. Distinctions in the reasons for smoking beginning of men and women have been analysed. Termination of smoking for women is more characteristic during the fertility periods; women have a greater elasticity of factors connected with self-estimation of health and quantitative family structure. Prevalence of smoking habits is higher in the cities because public opinion on harmful habits is more democratic there. The addictive character of cigarettes consumption has been confirmed empirically. Intensity of work and employment influences the quantity of consumed tobacco profoundly. We have obtained that smoking can be reduced through propaganda of a healthy life-style because physical activity restrains the use of tobacco and raises the probability of smoking termination.

1. INTRODUCTION

Due to Goskomstat data the population of RF spends regularly 4–5% of their income on tobacco and alcohol.[1] In 2001 Russians spent 1.46 bln $ on cigarettes (by estimations of Goskomstat RF), and according to IRG VS the expenses constituted 5–6 bln $.[1]

Smoking is number one enemy. 67% of men (it's three times more, than, for example, in the USA) and 30% of women smoke in Russia (Table A1). However it is not still clear, why they do smoke, understanding that they ruin their own health and of those around them exposed to passive smoking. One of the major tasks is to prevent young adults from desire to start smoking under the influence of their companions. Adults begin smoking insignificantly. Smoking starts very early before age of 11. Approximately 8–12% of schoolchildren of 7th–8th forms smoke regularly; in 9th–10th forms this figure constitutes 21–24%.[2]

It is necessary to foresee, what measures can result in stimulation of healthy lifestyle, and what consequences in Smoke-preventing State programs can be. It is not still clear how the change of prices on cigarettes influences tobacco consumption in Russia. It is necessary to understand the reasons inducing the beginning and the continuation of smoking. The extention of male smoking in the countries with the advanced market economy constitutes approximately 37%, in comparison with 60% in the former socialist countries of Central and Eastern Europe (Table A1). The most daily consumption per one smoker, 24 cigarettes a day, is observed in the countries with the advanced market economy, whereas in the former socialist European countries the appropriate figure was approximately 18.[3] It is not clear, whether it is possible to expect that, with the aging of smokers in post Soviet Russia and with their tolerance to nicotine, the daily consumption will also be increasing. Will the growth of incomes speed up this tendency or will it reduce smoking? The research of these questions at the micro level is urgent.

The economic aspects of tobacco consumption will be dealt with in the present research. The main goal of the research project is to analyse the determinants impacting the overcoming of the double-hurdle, *i.e.* the beginning of smoking and failure of smoking in Russia nowadays. It assumes the solution of the following main tasks:

- estimation of impact of socio-economic factors on initiation of tobacco consumption and smoking termination in view of immunity of those who never smoked or quitted smoking;
- to reveal individual and economic factors which influence the amount of tobacco consumption;
- to find out the gender and age peculiarities of tobacco consumption, especially for teenagers;

---

[1] On the data www.ej.ru/069/tema/03/index.

[2] From interview dr. Mikko Vienonena (on materials http://demoscope.ru).

[3] From http://www.sigarets.ru.

- to reveal the addictive character of smoking.

There is a variety of studies related to other aspects of smoking economy. The complete survey of available theoretical and empirical works is done by Chaloupka and Warner (1999).

The addictive nature of smoking is used in many recent studies on users' demand. Most often the researchers follow the rational model of addictive behavior offered by Becker and Murphy (1988).

The related questions of why people start smoking and why they quit are central to formulating and evaluating effective antismoking policy. Duration models, which are designed to measure the probability of transition between states, are the appropriate statistical tool for investigating the structural determinants of transition between addiction and non-addiction.

The main problem at construction of these models consists in sample censoring, *i.e.* individuals do not smoke either because they have never smoked, or because they have stopped smoking. This is the double-hurdle problem.

In this paper the starting hazard rate is defined as the probability that an individual will start smoking in a given time period, conditional on being a non-smoker continuously until then. The quitting hazard rate is the probability of quitting smoking in a given time period, conditional on being a smoker since the starting date. The two rates will respond differently to different stimuli, since the starting hazard reflects the point of view of nonsmokers, and the quitting hazard reflects the point of view of smokers.

Duration models possess several advantages over smoking participation models, for example in a participation rate model it is difficult to control for the effect of the duration of the habit.

It is some previous work in the economic literature that estimates the hazard rate of starting (quitting) smoking. Douglas (1998) investigated the determinants for the decisions to start and quit smoking in a rational addiction framework using the ordered probit split sample duration model with controlling nonsmokers, newly-started and quitted smokers, started and continued to smoke individuals. The price of cigarettes was a time-varying covariate. The log-logistic distribution was used for a model of smoking initiation and the Weibull distribution was used for a model of smoking termination. Douglas proved the significance of price in both models, though price effect in the model of smoking initiation is small. Jones (1989) estimated the double-hurdle probit model for attempt to quit smoking and its successful utilization. However, the important price variable has not been included into the model.

In their detailed paper Forster and Jones (2001) got tax elasticity equal to 0.16 for men and 0.08 for women for the age of smoking beginning, and –0.60 and –0.46 accordingly for the smoking duration. Forster and Jones described problems with tobacco consumption data and also econometric opportunities to test specification mistakes of duration models that we'll use in recent research.

Labeaga (1999) attempts to take into account a censoring problem. He uses a double-hurdle rational addiction model assuming that an individual consumes tobacco when he is a potential smoker (*i.e.* he likes to participate) and then he does not maximize utility at zero (*i.e.* he really participates).

Parametric and semi-parametric duration models to model multiple cessation attempts of young adults were estimates by Tauras (1999, 2001). The estimates indicate that price growth of cigarettes increases the probability of initial smoking cessation as well as subsequent cessation. The average price elasticity of cessation is 0.343. In addition, restrictions on smoking at private worksites and public places give rise to the probability of smoking cessation for young adults.

Beenstock and Rahav (2003) used long-term survivor models to research the initiation process into cannabis and hard drugs with deal of censoring individuals, *i.e.* who at time of data collection had not initiated drugs. The authors postulate that censored individuals come from two populations: a population that is "immune" to drugs, who will never use them no matter how long they live, and a population which is "susceptible" to drugs for whom it is a matter of time until they begin to use drugs. They are used parametric mixture model for immunity and initiation of cannabis that motivate to construction in this research composite model with dependence between quitting and consumption with exogeneity of smoking initiation and endogeneity of smoking consumption and quitting.

There are no traditions of econometrical analysis of tobacco consumption in Russia. Negative effects of smoking are established more often in available fragmentary researches.

2. INITIATION AND TERMINATION OF SMOKING: COX'S PROPORTIONAL HAZARD MODEL

We consider RLMS panel data allowing to expand essentially a framework of traditional statistical data analysis without being limited by analysis of correspondence tables and other standard tools. As the data make it possible to establish the date of beginning and termination of smoking, we use in this case the most adequate econometrical instrument of survival analysis for the "life-time data".

We use the respondents' information about the age of smoking initiation to define the calendar year of starting to smoke for those individuals who ever smoked. But the sample also contains the data on those individuals who never smoked. In a semi-parametric duration model such observations are interpreted as incomplete spells, and it is accepted that all individuals suffer failure at the initiation of smoking. These observations are classified as right-censored according to RLMS year of survey (Beenstock and Rahav, 2003).

At the initial stage the nonparametric analysis is used to analyse the factors impacting the hazard of smoking beginning (stopped). It is especially convenient to put forward hypotheses concerning the theoretical distribution of spells before the moment of termination. One of the results of these analyses is that, when estimating the hazard and survival functions, it allows to receive graphic representations of these functions in view of the censored data. Kaplan–Meier's product-limit estimator is the most widespread nonparametric method (Lancaster, 1990). The advantage of this estimation is that it allows to take into account the presence of censoring and of equal spells of observations in sample.

The main problem is to compare the curve survival functions for various strata/subgroups.[4] The similarity of curves can be appreciated with some criteria. We shall use log-rank statistics, as the most widespread criterion, to define whether to include a corresponding variable into the resulting model or not.

Parametrical and semi-parametrical models are used to estimate the joint effect of the factors impacting the smoking initiation/termination (Cox and Oakes, 1984; Lancaster, 1990). Cox's proportional hazard model is reliable and widely used in this case (Cox, 1972). Proportional hazard model is pertinent due to its simplicity and easy interpretation of the basic idea, *i.e.* the impact on the event duration corresponds to multiplication of hazard function to a constant multiplier which is defined by impact of exogenous factors:

$$\lambda(t|X,\beta)=\lambda_0(t)\varphi(X,\beta)\,, \qquad (1)$$

where X — a vector of explanatory variables, β — a vector of parameters subject to estimation, $\lambda_0(t)$ — baseline hazard function, which would correspond to resulting hazard function under the absence of influence of exogenous factors, *i.e.* the baseline hazard function presents the endogenous risk of the event termination.

Let's define individual specific risk function $\varphi(X, \beta)$ in the exponential form:

$$\lambda(t|X,\beta)=\lambda_0(t)\exp(\beta' X)\,. \qquad (1a)$$

One of the advantages of this method is that it does not need any specification form to a baseline hazard function (Cox, 1972; Cox and Oakes, 1984). However, one of the important assumptions to the model (1) is the proportionality of hazards, *i.e.* the relation of hazard functions for two members of totality should remain constant during the observation period. The regression parameter β_j of this specification (1a) produces a proportional effect of impact on a risk degree by absolute changes of a corresponding variable X_j:

$$\beta_j=\partial \ln\lambda(t,X)/\partial X_j\,.$$

Coefficient β_j also gives hazard elasticity on X_j (Jenkins, 2004). Estimations of parameters in Cox's model (1a) can be found with a maximum-likelihood method (Cox, 1972).

To build the smoking termination model, we use the smoking duration as a survival time measure and also use a sample of individuals who sometime smoked before and then either stopped smoking or continued to smoke. The last ones will be subject to the right-censoring according to RLMS year of survey.

Because of the differences in distributions of separate individuals, in spite of the account of pertinent exogenous variables, there can arise an unobservable heterogeneity which leads to biasing of

[4] Command *sts* in Stata 8 allows to find Kaplan–Meier's estimations and to test the hypothesis about absence of distinctions in survival functions by subgroups.

estimations and deceptive conclusions on the effect of explaining variables upon risk (Heckman and Singer, 1984). To account the unobservable heterogeneity in model (1a) we add random, specific for separate groups of individuals, variable g_i with gamma distribution:[5]

$$\lambda_{ij}\left(t, X_{ij}\left|g_i\right.\right) = \exp(\beta' X_{ij})\lambda_0(t) g_i = \exp(\beta' X_{ij} + v_i)\lambda_0(t)\,, \qquad \text{(1b)}$$

where i — a group heterogeneity index, j — a group observation index, g_i — unobservable heterogeneity, and $E(g_i) = 1$, $Var(g_i) < \infty$ and $v_i = \ln g_i$.

It is possible to test the importance of heterogeneity in model (1b) on the basis of dispersion estimations of v_i and its standard error. The choice between models (1b) and (1a) is also carried out with the help of likelihood ratio statistics having χ^2 distribution.

We use appropriate tests for checking up the constructed models to find out specification mistakes. We can easily control the conformity of shares for started (quitted) to smoke predicted according to the model and realized in a sample. To observe failures in sample, we use graphic representation of Cox–Snell's cumulative residuals (Klein and Moeschberger, 1997; Lancaster, 1990) to estimate the fitting of the model for those who refuse to smoke (continue to smoke per the model of smoking termination). Correctly fitted model should produce the Cox–Snell's cumulative residuals similar to the sample of standard exponential distributions. That's why, the graph of a nonparametric estimation of cumulative hazard function for such data should be on a line 45° from the origin of coordinates (a bisector of the first quadrant). The more the deviation from a bisector of the esteemed curve is, the worse the model is specified.

To test the basic assumption of Cox's model of risks proportionality, we use the scaled Schoenfeld residuals and also the test for separate variables and Grambsch and Therneau (1994) global test as well. Scaled Schoenfeld residuals are not dependent on time and have a zero expectation at a null hypothesis on proportional risks.[6] For alternative, when risks are disproportional, the residuals depend on time. The global test has χ^2 asymptotic distribution.

3. INFORMATION BASE OF RESEARCH, SAMPLE CONSTRUCTION AND CHARACTERISTIC OF VARIABLES

Samples were made on the basis of the 5^{th}–10^{th} waves (1994, 1995, 1996, 1998, 2000 and 2001) survey according to RLMS (Russian Longitudinal Monitoring Survey). About 11 thousand people in 4 thousand households were questioned. RLMS gives socio-economic characteristics to the population, including tobacco consumption and retrospective of smoking. We use the data of two

[5] Estimation of model with heterogeneity is executed by procedure *pgmhaz* in Stata 8 (Jenkins, 1997).

[6] We use the command *stphtest* with options in Stata 8.

types: household and individual. The merge of samples is carried out on the base of the individuals' identification numbers. The sub-samples of individuals (aged 14–65) with their family characteristics and appropriate set of basic investigated variables are analyzed. The first sample to study the smoking initiation and termination is formed as a pooled one for all individuals ever participating in survey. The second sample is formed as the panel for smoking individuals where we know the amount of consumed cigarettes.

3.1. Duration data

The RLMS database contains questions of the individual questionnaire, *e.g.* "Do you now smoke?" and "Have you ever smoked?", allowing to share all individuals into the sub-groups:

— never smoking;

— earlier smoking and stopped to smoke;

— smoking now.

In two last groups, using the basis of the following questions "Remember, please, when did you start smoking? How old were you then?", "How many years ago did you quit smoking?" and "Please remember, when did you start smoking? How old were you then?", it is possible to find out the time of smoking beginning and to define when the individuals have stopped to smoke. Thus, as for smokers, we precisely know the age of smoking beginning (calendar year) and the quantity of daily smoked cigarettes (question "About how many individual cigarettes or papyrosi do you usually smoke in a day?"); as for stopped smokers we learn the smoking period (calendar year of smoking termination); for non-smokers we get information that they never smoked before. At the same time we do not exactly know whether smoking individuals will stop smoking in the future or non-smokers will start smoking; therefore, for such individuals we have right-censored data (Fig. 1).

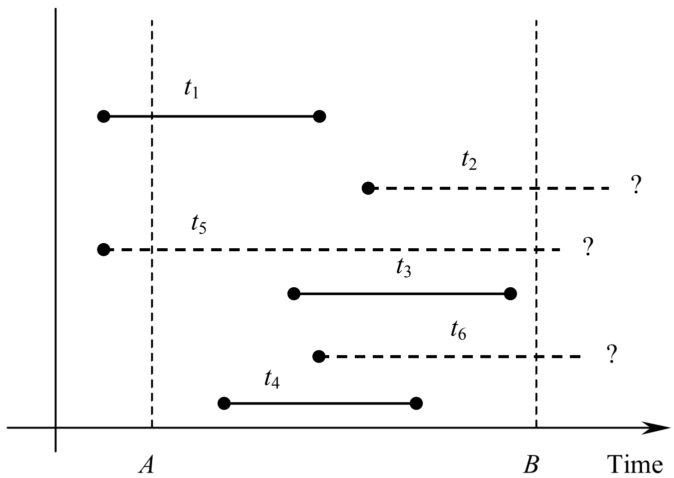

Fig. 1. Duration data.
A — Beginning of study period (1994); B — End of study period (2001); t_2, t_5, t_6 — Right-censored spells; t_1, t_3, t_4 — Completed spells.

The sample for those started to smoke contains 4798 individuals (including 24% of men), from which 399 persons have started to smoke (Table A2). The sample for stopped smokers contains 1926 individuals (including 77% of men) from which 138 persons have stopped to smoke. On the average young men start to smoke at 16.5 y.o. and girls a little bit later — at 19.7. Stopped smokers used cigarettes about 12 years running, and women quit to smoke, on the average, after 5 years of smoking, *i.e.* at the age of about 25. Men smoke more intensive, on the average, about 15 cigarettes a day (*i.e.* 3/4 pack of cigarettes), women smoke less, on the average, about 9 cigarettes a day.

The structure of a year-to-year sample, based on RLMS survey, is submitted in Table A3. It is obvious, that the shares of men and women consuming nicotine have been leveled by 7% in the period from 1994 up to 2001 because of the increased use of cigarettes by women; moreover, 4% that of specified are achieved due to involving into smoking the young aged up to 30.

The information on distribution of average smoking duration depending upon age is submitted in Table A4. It is worthy to mention that increase of average smoking duration goes in parallel with the increase of smoker's age. The additional data analysis allows to conclude that children start to smoke before their parents. The accretion of smoking duration is gained after 30 y.o., it means that considerable part of individuals stop to smoke at the age before 30.

After answering the question "Do you now smoke?" which was put to the same individuals at different RLMS waves, there were defined the individuals who smoked, quitted smoking and then started to smoke again — totally 571 persons, 64% of which were men. 58% of individuals, who tryied to stop smoking, the break period constituted no more than 2 years. Distribution of the specified individuals by cohorts is the following:

Categories	Under 20	21–30	31–40	41–50	Older 50
Frequency, %	23	26	25	15	11

It is impossible to establish precise year of smoking beginning or smoking termination for all of them (because RLMS survey was not carried out in 1997 and 1999); not all of them have a full set of explaining variables. Therefore, we take account of 40 individuals as beginners in the model of smoking initiation and 18 individuals as stopped smoking in the model of the smoking termination out of mentioned 571 persons.

3.2. Price data

The information on prices for one pack (20 cigarettes) can be obtained from the answers to the following questions of the family questionnaire: "Did your family buy in the last 7 days tobacco products?", "How much packs?" and "How many rubles in all did you pay?". However, such data can be biased because the questions have been frequently responded by non-smokers, and there are also many observations viewed in RLMS with absent answers to the specified questions. Data is missing for those years when the survey had not been carried out.

The other source of price information is Goskomstat data in RF regions for the period beginning from 1995. We used the Goskomstat indices on average consumer prices for cigarettes of domestic manufacturing for 1 pack in some cities of Russian Federation (computed on the basis of retail trade data) per every month of the RLMS survey or only per December, if the survey had not been carried out at that year.

Pricegks variable has been generated in the following way. 38 primary sampling units (*psu*) of Russian Federation territorial divisions have been determined. The cigarettes price in the period of 1995–2001 was defined as the Goskomstat price for cigarettes in the cities corresponding to *psu*. To extend the sample we used the All-Russia cigarette costs indices published by Goskomstat (price in December related to December price of a previous year) and calculated the prices for the mentioned 38 units consistently, in the period of 1993–1994, by division the regional price of the current year into the consumer cigarette costs index. Such a definition of price in 1993–1994, undoubtedly, gives an error; however, it allows to take into account the formed regional price differentiation and, to our mind, it is quite a good approximation. Besides, before 1993, when individuals began to smoke, data on prices were unavailable. This fact essentially changes the sample structure to research the initiation of smoking; the number of men started to smoke before 1993 has reduced significantly. However, there is no way to by-pass the specified difficulties.

Later cigarette prices were deflated by the cost of consumer's basket in regions[7] (according to Goskomstat). Real price values are presented in Table A5.

The comparison of RLMS and Goskomstat prices shows that RLMS prices are more often underestimated and the average difference constitutes about 20%, that is satisfactory.

The cigarettes price variable is a relevant political tool to regulate tobacco market. That's why our results can be considered as elasticities influenced by a relative change of a price level from the beginning up to the termination of smoking.

The retail price on cigarettes includes a selling price with excise-duties, a trading extra charge to the selling price, a value-added tax and transport costs. The value of the trading extra charge is approximately identical in the regions. Rates of excise-duties and taxes are adjusted at the federal level. The excise-duty rate on tobacco products depends on a type of product: tobacco, cigars, filtered cigarettes, unfiltered cigarettes; and this rate is not influenced by the contents of nicotine in a product. In view of working in regions of Russia more than 60 tobacco factories, but transport costs do not play any essential role in forming of retail prices. Thus, it is not still clear, what is the reason of inter regional variations in prices on tobacco products. Probably, these price variations appear due to the differences in consumption in the regions: in those regions with a higher standard of living people smoke high quality and expensive cigarettes, and in the regions with a low standard of

[7] Because the calculation methodology of consumer's basket has changed in 2000 we used values of consumer's baskets in regions on the end of 2001. After deflating the prices on these baskets, we used a consumer price index for reduction of prices to necessary years. The variation between years in the regional baskets ratio is small by assuming, and it can be neglected.

living people smoke cheap cigarettes. These price variations reflect rather differences in incomes in the regions, and it should be taken into account while interpreting the elasticity. We shall generate proxy variables of regional prices for cheap and expensive cigarettes taking into consideration the preferences in cigarettes consumption across regions. Obviously, it is possible to assume, that poor households consume only cheap cigarettes and rich households consume, as a rule, only expensive cigarettes. Then we'll find bottom (2^{nd}) and top (8^{th}) deciles of household distributions according to incomes and average RLMS prices of tobacco calculated for these households by *psu*. We'll receive the cigarettes price *pricel* for 20% of poor households and the cigarettes price *priceu* for 20% of wealthy households. The received price variables are deflated by the cost of consumer's basket in the regions (Fig. A1).

Thus, we have three price variables: determined according to Goskomstat — *pricegks*, and according to RLMS data — *pricel* and *priceu*.

Goskomstat cigarettes real price (1993–2001) constitutes average about 0.35 rbl per a pack, RLMS price (1994–2001) — about 0.27 rbl per a pack for cheap cigarettes and 0.36 rbl per a pack for expensive ones. Let's notice that beginners consume more expensive cigarettes, than those terminated to smoke: the real average price (on Goskomstat) makes, accordingly, 0.48 and 0.31 rbl per a pack. Hence, price elasticity should be higher for smoking termination in comparison with smoking initiation.

Let's note that expenses on tobacco products constitute average about 5% of incomes for 20% poor RLMS households and about 0.8% for 20% wealthy RLMS households in the period of 1994–2001.

As far as we know the years of beginning and termination of smoking, it allows us to mix RLMS data with the prices for cigarettes data. The correspondence between the prices and RLMS data was established due to the year of beginning and termination of smoking; as for non-smokers (for smokers — in termination model) — it was established according to the last wave of survey in which the individual participated. The individuals, who began to smoke (or stopped smoking — in termination model) before 1994, were removed from the pooled sample because the price for cigarettes was unknown to them at the moment of smoking beginning (termination of smoking). Thus, the sample contains only those individuals, who began to smoke not later than 1994, and those who did not begin to smoke and for whom the price of cigarettes in the appropriate year has been defined. The sample for smoking termination model contains, accordingly, only those individuals, who have stopped to smoke not earlier than 1994 or continued to smoke.

3.3. Explaining variables

There exist potential problems associated with predicting of past behaviour as a function of individual characteristics that are measured at the time of RLMS survey. That's why we use a parsimonious set of exogenous covariates for starting and quitting models. We try, as far as possible, to use covariates that were exogenously determined prior to the individual's starting or quitting decision. Therefore, we avoid covariates, such as health and past smoking status that may be prone to unobservable heterogeneity bias.

Let's mention, that RLMS data, as well as any retrospective ones, contain errors of two types: the respondents do not always accurately remember, when they have begun to smoke; and, as they are not sure in questionnaire anonymity, they can not exactly answer to the questions on the amount of smoked cigarettes. However, as it is noted in Beenstock and Rahav (2003), these errors can not strongly influence the size of model estimations. To estimate this displacement according to Tunali and Pritchett (1997),[8] the data are being transformed the way, when the duration variable is measured in a calendar time period both for the age of beginning and the age of smoking termination. We shall apply this method of measurement of time duration to our research.

Sets of exogenous variables in models of smoking initiation and smoking termination are differ, except for individual characteristics (gender, marital status, self-estimation of health during the period previous to the termination or initiation of smoking) and the account of cohort effect and time trend. We consider that the beginning of smoking in the greater degree is caused by urbanization, in particular, more democratic norms of public morals in city in comparison with countryside. Also the beginning of smoking essentially depends on number of members of family (proxy for the income and household structure). At the same time, in our opinion, the termination of smoking is impacted with duration of smoking and physical activity of the individual, and also his profession. We assume that physical exercises are poorly compatible with smoking. Quitting of smoking with "white collars", in our opinion, is more probable as compared to individuals of working professions. Addictive nature of smoking comes out, in particular, with expectation of a negative influence of smoking duration on tobacco termination. Obviously, the smokers with higher incomes are less often inclined to stop a harmful habit.

Distribution of smoking duration for stopped individuals is shown in Fig. A2, from which it is obvious that about 35% of smokers stopped to smoke per the first 5 years after the beginning of smoking. The computations show that among quitted-smoking teenagers about 90% have stopped per the first 3 years after the beginning of smoking (Fig. A3).

The distribution of average quantity of daily smoked cigarettes, depending on sex and cohort, is shown in Fig. A4. The average quantity of daily smoked cigarettes is increasing steadily for men with a peak at the age of 46–50. As for women, age dynamics of tobacco consumption increases unstably with a peak at the same age of 46–50.

4. THE ANALYSIS OF FACTORS IMPACTING THE INITIATION OF SMOKING

Survival and hazard functions are estimated by the Kaplan–Meier's procedure for duration of smoking abstention. Survival function is submitted in Fig. A5 for full sample, and in Fig. A7 for

[8] Tunali, I. and J. Pritchett (1997) Cox regression with alternative concepts of waiting time: the New Orleans yellow fever epidemic of 1853, *Journal of Applied Econometrics* **12**, 1–25.

men and women separately. Also hazard functions are submitted in Figs A6 and A8 for full sample and by gender.

Survival function decreases monotonously, beginning from the age 16 quickly enough. Hazard function is monotonously growing with strongly pronounced local maximum for the age about 20 without sharp peaks on the whole interval of definition. The hazard function for women is less convex than the analogue one for men. Thus, smoking beginning starts at the age before 20; young men start smoking more often than girls.

We also investigated in detail the dependence of hazard of smoking beginning upon birth cohort. Figs A9, A11 and A13 present the Kaplan–Meier's survival functions for cohorts: before 30 y.o., from 31 to 50 and after 50. Figs A10, A12 and A14 present the appropriate hazard functions. We can see that the survival function's decrease begins after 10 years of non-smoking for youth, for adults — after 20 years, for old men — after 40 years. Probably, such thresholds with steps of 5 or 10 years are caused by a "heap effect", because the respondents have approximated the number of years in their answers to the nearest numbers divisible by 5 or 10. In Forster and Jones (2001) the ways of model correction are discussed in view of the specified effect[9] and the analysis of estimations' sensitivity for different ways is carried out; it shows that the changes in estimations are insignificant. We used the ad hoc approach of addition of dummy variables for "heaping" observations and also have not received essential changes in estimations. The appropriate distinctions of smoking beginning hazard are seen in a Figs A10, A12 and A14. The bell-shaped curve of hazard function, both for boys and girls, with explicit maximum at 17, corresponds to age cohort before 30. The hazard function curve for adult women is relatively gently sloping and is almost constant for women after 50. The hazard function curve for men aged between 31 and 50 has its maximum at 28 that is, probably, due to the life cycle: a man at this age actively provides for his family, makes a career and when being under stress he starts to smoke. The hazard function curve for men after 50 varies and has two appreciable peaks at the age after 40 due to the crisis of mature man's age. Thus, it is necessary to take into account the cohort effect in the model.

We shall include the following factors into the set of explaining variables of vector X: gender (*female*), marital status (*marst*), urbanization (*urban*), deflated price of a cigarette pack (*pricel* and *priceu*), number of members of households (*numhh*), dummies variables for the cohort effect and dummies variables for years of RLMS survey.

It is obviously necessary to take into account the social norms in the model specification, in particular, the prevalence of smoking among relevant groups. However, the RLMS questionnaire does not contain corresponding questions on where and with whom the individual smokes. In this case, taking into account Russian traditions of joint alcoholic drinking, the indicator of drinking alcohol becomes a good instrument. Since alcohol consumption is the endogenous variable for smoking we'll use the *selfhealth* proxy variable of health self-estimation of an individual. To exclude the en-

[9] By article: Torelli, N. and U. Trivellato (1993) Modelling inaccuracies in job-search duration data, *Journal of Econometrics* **59**, 187–211.

dogeneity between *selfhealth* and initiation of smoking, we shall use a lagged variable related to smoking initiation.

The sample contains 42% of married men and 54% of married women. 2/3 of individuals live in a city. Men characterize their health as good, women — as average (Table A2).

Table A6 presents the values of logarithmic rank criterion of checking the survival functions equality after Kaplan–Meier for variables included into the smoking initiation model. The results allow to judge about a significant discriminative ability of all variables.

Estimations of parameters of Cox's proportional hazard model (1a) of non-smoking duration for full sample and separately by gender are submitted in Table 1. The model of smoking initiation is also constructed with the excluded lagged variable *selfhealth* to increase the sample volume where Goskomstat prices are used — *pricegks*. Results (Table A7) essentially do not differ from those obtained in Table 1.

Graphics of Cox–Snell's cumulative residuals (Fig. A15) allow making a conclusion that models provide a satisfactory quality of forecasting, since the residuals are arranged near on a line 45°, except for the end of the interval. The predicted share of started smoking individuals is presented in the bottom of Table 1. It is equal to 0.13 for a common model that is close to the actually observed value 0.08. Testing of proportional hazards hypothesis on the basis of scaled Schoenfeld residuals, both for global model and for variables, has given results submitted in Table A8. The assumption on hazards proportionality has been proved. We estimated models (full sampling and separately by gender) assuming the possibility of unobservable heterogeneity with gamma distribution of form (1b). The value of dispersion ν_i gamma distribution, relating to its standard error, has not exceeded a value 0.01, that allows to conclude on the insignificance of the unobservable heterogeneity for our data.

Calculations show high price elasticity of hazard and differences in population response to the increase of prices on cheap and expensive cigarettes. We assume that people in rich households smoke good and expensive cigarettes, and in poor households they smoke cheap cigarettes. With increasing the real price on cheap cigarettes by 1 kop smoking probability increases by 0.03 (≈exp(0.02781)) in a short time period. The explanation of such paradoxical fact is the further decline in living standards in parallel with the increase of prices which results in stress and harmful habits. Increase of real prices on expensive cigarettes by 1 kop. per a pack results in reduction of smoking beginning probability on 0.025; men's price responsivity is higher — the probability is reduced on 0.03, as compared to the women's one — on 0.02. According to the estimations, the risk of smoking beginning for women is 67% less. At the same time, a proxy variable of health self-estimation in the period previous to smoking beginning (*selfhealth*) is a significant factor influencing the initiation of smoking, and this factor is significant for women by raising their smoking probability on 0.14, if in the previous period the health became worse.

Individuals living in a city are more inclined (on 30%) in smoking initiation as compared to those living in a countryside. This factor is more important for women, than for men. It is because public opinion in a countryside still remains the influential factor and hinders smoking distribution as it considers smokers, especially women, to be violaters of social norms. At the same time public opinion in a city concerning harmful habits is indulgent and democratic.

Table 1. Cox's model results for smoking initiation.

Variables	Full sample	Men	Women
Female	-1.118^{***} (0.104)	–	–
Selfhealth	0.127^{*} (0.072)	0.023 (0.087)	0.338^{***} (0.125)
Marst	–0.033 (0.148)	0.012 (0.224)	–0.129 (0.198)
Urban	0.260^{**} (0.112)	0.234^{*} (0.139)	0.297^{*} (0.168)
Pricel	2.781^{***} (0.389)	2.779^{***} (0.438)	2.784^{***} (0.798)
Priceu	-2.619^{***} (0.385)	-2.965^{***} (0.446)	-2.054^{***} (0.716)
Numhh	-0.096^{***} (0.037)	-0.080^{*} (0.047)	-0.133^{**} (0.062)
30<Age≤50	-2.479^{***} (0.241)	-2.415^{***} (0.390)	-2.579^{***} (0.303)
Age >50	-3.939^{***} (0.462)	-3.540^{***}(0.731)	-4.322^{***} (0.597)
1995	0.546 (0.627)	0.315 (0.921)	0.871 (0.863)
1996	0.686 (0.561)	0.531 (0.780)	0.994 (0.809)
1998	0.914^{*} (0.541)	0.758 (0.750)	1.239 (0.783)
2000	0.738 (0.551)	0.783 (0.760)	0.697 (0.819)
2001	0.858^{*} (0.521)	0.818 (0.729)	0.983 (0.750)
No. of subjects	4798	1129	3669
No. of failures	399	250	149
Log likelihood	–2870.14	–1556.47	–1045.65
χ^2	527.10	152.10	180.46
Predicted proportion of starters	0.13	0.39	0.29
Observed proportion of starters	0.08	0.22	0.04

The values of coefficients are presented in the table. Breslow method for ties. The robust standard errors adjusted for clustering on individuals are in brackets.

Sample contains individuals started to smoke not earlier than 1994.

Significance of coefficients: * — 10% level, ** — 5% level, *** — 1% level.

Marital status is not a significant variable for initiation of smoking in all specifications of the model. On the other hand, the variable of quantity of household members has been included into the model, allowing to affirm that with the increase of a household "size" on 1 person the risk of smoking beginning for its members decreases 9% for full sample and 11% for women that is caused, first of all, by giving birth to children.

The generations' effect is significant. Born in 70's are more inclined in smoking initiation, than those born in 50's. Time-dummy variables are significant and show that risk of smoking beginning increases significantly in 1998 and later in 2001; it allows to make a conclusion on a high probability of smoking initiation probability in the first decade of 2000. Obviously, such situation appears due to changed social and economic conditions of life in comparison with those the middle of 90's.

We have constructed Cox's proportional hazard models especially for the young up to 20 years of age inclusively, because teenagering is a critical period for smoking beginning. The results are submitted in Table 2.

Table 2. Cox's model results for beginners under age of 20.

Variables	Full sample	Men	Women
Female	-1.121^{***} (0.141)	–	–
Selfhealth	0.179^{*} (0.096)	0.069 (0.116)	0.411^{*} (0.182)
Marst	0.294 (0.358)	0.062 (0.663)	0.309 (0.448)
Urban	0.097 (0.139)	0.043 (0.163)	0.270 (0.265)
Pricel	3.182^{***} (0.477)	3.208^{***} (0.548)	3.234^{***} (1.104)
Priceu	-2.740^{***} (0.489)	-2.984^{***} (0.597)	-2.296^{**} (0.925)
1995	–0.553 (0.714)	–0.324 (0.939)	–0.617 (1.135)
1996	–0.537 (0.607)	–0.456 (0.820)	–0.441 (0.916)
1998	–0.726 (0.570)	–0.698 (0.772)	–0.451 (0.881)
2000	-1.226^{**} (0.584)	–0.874 (0.792)	-1.810^{**} (0.895)
2001	-1.279^{**} (0.537)	–1.031 (0.745)	-1.611^{**} (0.781)
No. of subjects	986	434	552
No. of failures	226	155	71
Log likelihood	–1427.40	–870.21	–407.71
χ^2	192.71	72.38	61.14
Predicted proportion of starters	0.06	0.05	0.34
Observed proportion of starters	0.23	0.36	0.13

The values of coefficients are presented in the table. Breslow method for ties. The robust standard errors adjusted for clustering on individuals are in brackets.

Sample contains individuals started to smoke not earlier than 1994.

Significance of coefficients: * — 10% level, ** — 5% level, *** — 1% level.

The results of testing on proportionality are submitted in Table A9. The predicted and observed proportions for beginners are presented in the bottom of Table 2. As a whole, the error is rather great (probably because of a small sample size), that is due to poor prognostic properties of the obtained models. The same conclusion follows from the Cox–Snell's cumulative residuals graphics.

Nevertheless, the important variables *pricel* и *priceu*, which are the "responses" to the economic policy of the authorities are significant. The increase in prices of expensive cigarettes by 1 real kop. reduces the probability of smoking beginning on 0.027; as for young men, the sensitivity to price changes is more — on 0.03, as for girls it is less — 0.023.

The comparison of outcomes from the model for full sample (Table 1) demonstrates that elasticity of price factor for the young is hardly higher than for all individuals. The increase in prices of cheap cigarettes raises the risk of smoking beginning in a short-time period almost 1.03 times; tit indicates a social problem, *viz.* decline of life standards together with growth of real prices leads to a wide practice of a harmful habit of smoking among teenagers. Let's mark a positive impact of health decline on smoking initiation with the girls. Factors of urbanization and marital status are not significant.

5. THE ANALYSIS OF FACTORS IMPACTING THE TERMINATION OF SMOKING

Survival and hazard functions are estimated by Kaplan–Meier's procedure for smoking duration of individuals who smoke. Survival function for full sample is submitted in Fig. A16, for men and women separately in Fig. A18. Hazard functions for full sample and by gender are submitted in Figs A17 and A19. The survival function monotonously descends that is typical for both men and women. The hazard function for full sample oscillates with two local maximums — for duration of smoking per 10 and 40 years that correspond to the ages of 27 and 57. We note the different dynamics in hazard functions for men and women. The hazard function for women has a strongly pronounced oscillatory character with spikes for 8, 30 and 45 years of smoking duration, and decays for 23 and 38 years. Its values on the whole range of definition are practically larger than those for men. Hazard function for men monotonously increases up to the duration of smoking per 40 years and then begins to decay. We investigated in detail the dependence of hazard of smoking cessation upon birth cohort. The Kaplan–Meier's survival functions for cohorts before 30, from 31 till 50 and after 50 are submitted in a Figs A20, A22 and A24. The appropriate hazard functions are presented in Figs A21, A23 and A25. The risk to quit smoking is the highest possible for girls having 5 and 13 years of smoking experience and 5 years for young men. At mature age the risk to stop smoking has a peak for women with 9 and 25 years of smoking experience and for men with 24 years of smoking duration. For elder age the risk to stop smoking has a peak for men with 40 years of smoking experience and for women with 30 and 45 years of smoking. Thus, women very likely quit smoking at the ages about 24, 30 and 45. Smoking termination at 24 and 30 y.o. is connected with fertility periods and birth of the first/second child; at the age near 45 — with physiological changes of organisms. Men mostly quit smoking at the age of 22 and 41. In the first case it is the age of getting education, job-searching and marriage that result in changing of habits. At mature age men, obviously, stop to smoke because of problems with their health.

The duration of use of cigarettes and tobacco, apparently, is one of the factors influencing the determination to stop smoking. Therefore, we shall include a variable *durationcat* into the structure of vector of explanatory variables X and also the following factors: gender (*female*), self-estimation of health lagged as compared to termination of smoking event (*selfhealth*), marital status (*marst*), prices of 1 pack of cigarettes (*pricel* and *priceu*), physical activity (*sport*), professional dummy variable (*prof*), logarithm of the individual income for last month (*lnincome*), dummy variables for the cohort effect and dummy variables for years of RLMS.

On the whole, the sample presents individuals who estimate their health levels as average ones (Table A2) and do exercises less than 3 times a week. Women mostly have work at offices than men. Average duration of smoking of the individuals in sample is equal to 19 years for men and 12 years for women.

Table A6 submits the values of log-rank test of Kaplan–Meier's survival functions equality for variables included into the model of smoking termination. The results allow to judge the significant discriminant capacity of all variables except variable *marst*.

Estimations of Cox's proportional hazard models (1a) parameters are submitted in Table 3 for smoking termination in full sample and separately by gender. The model of smoking termination with Goskomstat prices — *pricegks* is also constructed. Results (Table A10) do not essentially differ from those submitted in Table 3, except for the price variable *pricegks*, having a negative sign and being significant in the model.

Graphics of Cox–Snell's cumulative residuals (Fig. A26) allow to conclude that model for women provide low properties of specification. The full model is satisfactory-specified. The predicted proportion for stopped smokers is given at the bottom of Table 3. For a common model it equals 0.03, that actually does not coincide with the observed value 0.07. However, testing of the hazards proportionality hypothesis, using the basis of scaled Schoenfeld residuals for global model and for variables, has given results submitted in Table A11, that demonstrate proportionality of explaining variables.

We estimated the models (in full sampling and separately by gender) to assume a possibility of unobservable heterogeneity with gamma distribution in view (1b). The value of dispersion ν_i gamma distribution, related to its standard error, has not exceeded a value 0.01; it allows to conclude that the unobservable heterogeneity is not significant.

Results of modelling show that individuals, consuming cheap cigarettes, have a positive price elasticity of smoking termination probability: increase of real price by 1 kop. per a pack of cheap cigarettes results in rise of probability to quit smoking during a short time-period by 0.02. At the same time, individuals with high living standards have an asymmetric response to price rises, individuals seldom quit smoking. It's paradoxical at first sight. At the same time, this fact testifies that price on cigarettes for people with high incomes is secondary to a smoking habit. In connection with a smoking habit the current price of cigarettes, probably, influences not so much the smoking termination as the displacement in demand for expensive cigarettes is connected with a high level of incomes. Thus, price elasticity of smoking termination is asymmetric by sorts of cigarettes — it is positive for individuals smoking cheap cigarettes and negative for those smoking expensive cigarettes.

Table 3. Cox's model results for quitted smoking.

Variables	Full sample	Men	Women
Female	0.272 (0.198)	–	–
Selfhealth	0.088 (0.159)	0.279 (0.195)	–0.251 (0.209)
Marst	0.115 (0.173)	–0.114 (0.221)	0.258 (0.258)
Sport	0.079 (0.069)	0.178** (0.075)	–0.230 (0.176)
Durationcat	–0.626*** (0.108)	–0.671*** (0.147)	–0.659*** (0.183)
Pricel	2.198** (0.762)	1.250 (1.003)	3.156** (1.248)
Priceu	–3.191*** (0.629)	–2.028*** (0.771)	–5.232*** (1.064)
Prof	0.624*** (0.185)	0.609** (0.247)	0.646** (0.283)
Lnincome	–0.022 (0.021)	–0.003 (0.027)	–0.059* (0.033)
30<Age≤50	0.375* (0.216)	0.703** (0.306)	0.038 (0.354)
Age >50	1.294*** (0.442)	1.706*** (0.582)	0.990 (0.744)
1995	–0.786 (0.503)	–0.909 (0.569)	–0.372 (1.219)
1996	–1.098** (0.503)	–1.017* (0.623)	–0.684 (0.958)
1998	–0.960** (0.398)	–1.349** (0.541)	0.126 (0.722)
2000	–1.351*** (0.419)	–1.453** (0.579)	–0.726 (0.738)
2001	–1.532*** (0.414)	–2.202*** (0.567)	–0.334 (0.719)
No. of subjects	1926	1492	434
No. of failures	138	82	56
Log likelihood	–892.75	–514.97	–273.04
χ^2	163.67	112.66	66.49
Predicted proportion of stopped	0.03	0.03	0.03
Observed proportion of stopped	0.07	0.05	0.13

The values of coefficients are presented in the table. Breslow method for ties. The robust standard errors adjusted for clustering on individuals are in brackets.

Sample contains individuals stopped to smoke not earlier than 1994.

Significance of coefficients: * — 10% level, ** — 5% level, *** — 1% level.

Rich women stop harmful habit of smoking less often. The factor of physical activity for men significantly influences the negative attitude to smoking and raises the risk of smoking termination by 19%. Hence, propagation of physical activity and sports is one of the ways to reduce smoking.

The hazard of smoking failure for service, craft and related trades workers, plant and machine operators and assembly workers, unskilled and army men is less as compared to legislators, top managers, officials, technicians and associate professionals and clerks. Difficulties of work, thus, influence the failure of unhealthy habits for men and for women to less extent.

The major factor causing continuation of smoking is the experience of smoking. This fact indirectly confirms addictive character of smoking.

Individuals born before 50's years of the previous century are more inclined to stop smoking.

Dummy variables for years of RLMS survey are significant in the model; risk of smoking termination is less in 2000. In combination with growth hazard of smoking initiation in the same period, we can come to the conclusion on spreading of this habit at the beginning of this century, that is consistent with the results of sample descriptive statistics: from 1994 up to 2001 the share of smokers has increased 3.6%, including 2% in 2000–2001.

6. MODELLING THE AMOUNT OF CONSUMED CIGARETTES FOR SMOKERS

Let's take advantage of the formed panel for individuals who smoked or continue smoking and who know exactly the quantity of consumed cigarettes.

According to RLMS data, the average quantity of daily smoked cigarettes in sample increased steadily from 14.52 (in 1994) up to 15.69 (in 2001), moreover, the growth for men constituted from 15.75 up to 17.31 and for women from 7.96 up to 10.20. The increment for women is almost 1.5 times more. The standard deviation of smoked cigarettes quantity for men is a little bit more than for women (from 7.16 up to 8.12 and from 6.04 up to 6.13 accordingly). Thus, it is possible to assert that, according to the received data, men smoke almost 2 times more than women.

More than 85% of smokers smoke filtered and unfiltered cigarettes, and also more than 90% of women smoke filtered cigarettes.

Let's in addition use the follow-up variables: *hours* — average duration of a working day for employed workers, hours; *age*, *agesq* — age of the individual and square of age accordingly; *sort* — kind of tobacco (1 — papyrosi, 2 — filtered cigarettes, 3 — unfiltered cigarettes, 4 — self-rolled cigarettes, 5 — pipe); *empl* — 1 for employed and 0 for unemployed, and also dummy variables for professional and educational categories.

We'll investigate the dependence of a log for smoked per day quantity upon variable factors on the basis of panel data structure and sample censoring for non-smokers. Therefore we use tobit model with random effects in errors (values of likelihood ratio test are satisfactory).

The outcomes of estimation are presented in Table 4 and in Table A12 with Goskomstat price variable. Because the relevant for us variable *sport* can not be constructed for the survey in 1994, the estimations are adduced for two cases with discharging of models for the employed in each case.

Table 4. Tobit random-effects regression for logarithm of amount of smoked cigarettes.

Variables	RLMS panel 1995–2001		RLMS panel 1994–2001	
	Full sample	For employment	Full sample	For employment
Const	–1.096*** (0.097)	–1.132*** (0.120)	–1.118*** (0.106)	–1.086*** (0.109)
Marst	–0.070*** (0.021)	–0.066*** (0.022)	–0.066*** (0.026)	–0.056*** (0.021)
Female	–0.876*** (0.027)	–0.848*** (0.029)	–0.909*** (0.026)	–0.878*** (0.027)
Age	0.016*** (0.006)	0.014** (0.006)	0.021*** (0.005)	0.018*** (0.006)
Agesq	–0.0003*** (0.00007)	–0.0002*** (0.00008)	–0.0003*** (0.00007)	–0.0003*** (0.00007)
Elementary Education	–0.029 (0.031)	–0.037 (0.032)	–0.009 (0.028)	–0.013 (0.029)
PTU with Secondary Education	0.033 (0.025)	0.042* (0.025)	0.036*(0.022)	0.037 (0.023)
Vocational Education	–0.012 (0.028)	–0.018 (0.029)	–0.011 (0.026)	–0.016 (0.027)
Higher Education	–0.061* (0.033)	–0.068** (0.035)	–0.051* (0.031)	–0.058* (0.032)
Legislators, Senior Managers, Officials	0.158*** (0.043)	0.198*** (0.045)	0.154*** (0.040)	0.191*** (0.042)
Professionals	–0.049 (0.039)	–0.032 (0.040)	–0.038 (0.035)	–0.022 (0.037)
Clerks	0.011 (0.052)	0.007 (0.054)	0.030 (0.048)	0.020 (0.050)
Service Workers&Market Workers	0.168*** (0.040)	0.126*** (0.042)	0.185*** (0.037)	0.164*** (0.039)
Craft&Related Trades	–0.011 (0.035)	–0.012 (0.037)	0.002 (0.033)	0.006 (0.034)
Plant&Machine Operators&Assemblers	0.031 (0.035)	0.006 (0.037)	0.047 (0.033)	0.028 (0.034)
Unskilled	0.005 (0.038)	–0.024 (0.039)	0.038 (0.035)	0.019 (0.037)
Army	0.232*** (0.079)	0.234***(0.084)	0.199***(0.072)	0.178** (0.076)
Sort	1.339***(0.010)	1.350*** (0.011)	1.310*** (0.009)	1.325*** (0.009)
Pricel	0.100 (0.062)	0.117 (0.076)	0.130 (0.098)	0.149 (0.101)
Priceu	0.214***(0.049)	0.194*** (0.053)	0.170*** (0.045)	0.157*** (0.047)
Sport	–0.019** (0.008)	–0.023*** (0.009)	–	–
Hours	–	0.014*** (0.002)	–	0.013*** (0.002)
Empl	0.047* (0.027)	–	0.062** (0.025)	–

Variables	RLMS panel 1995–2001		RLMS panel 1994–2001	
	Full sample	For employment	Full sample	For employment
1995	0.133^{***}(0.024)	0.138^{***}(0.026)	0.093^{***} (0.022)	0.033 (0.025)
1996	0.165^{***}(0.025)	0.167^{***}(0.026)	0.134^{***} (0.022)	0.075^{***} (0.025)
1998	–	–	–0.024 (0.023)	-0.089^{***} (0.026)
2000	0.194^{***} (0.022)	0.180^{***} (0.023)	0.169^{***} (0.022)	0.094^{***} (0.025)
2001	0.291^{***} (0.022)	0.288^{***} (0.023)	0.268^{***} (0.022)	0.203^{***} (0.025)
Log likelihood	–12229.58	–11016.95	–14874.26	–13370.83
Wald χ^2	22917.65	21216.48	26715.86	24719.67
Likelihood-ratio test of sigma_u=0, χ^2	2349.12	1918.32	3080.86	2581.44
sigma_u	0.628	0.617	0.631	0.619
sigma_e	0.617	0.616	0.627	0.628
rho	0.509	0.500	0.503	0.494
Observation:				
uncensored	8484	7652	10277	9236
left-censored	13518	12006	16677	14772

Standard errors are in brackets. Significance of coefficients: * — 10% level, ** — 5% level, *** — 1% level. Base categories: Secondary Education, Technicians and Associate Professionals.

We received that the dependence of a log for smoked cigarettes upon age has a quadratic relation with the extremum point at the age about 29, after which the quantity of smoked cigarettes begins to reduce. Women and married individuals smoke less. University education also positively affects the reduction of tobacco consumption.

The alcohol consumption variable is excluded as endogenous. Physical activity is a constraining factor to use tobacco, that allows once again to confirm a conclusion on a benefit of propagation of a healthy way of life and also of physical training lessons and exercises.

Changes in the structure of tobacco consumption, *i.e.* the increase of a proportion for unfiltered, self-rolled cigarettes and pipe smokers, raise a log of smoked cigarettes quantity in 1.3 times.

Employed individuals smoke more averagely 5% more as compared to unemployed ones, the smoked quantity is proportional to jobs difficulties, *viz.* to the quantity of working hours.

As to professional categories subject to a smoking habit, it is possible to approve that smoking is more popular with (as compared to technicians and associate professionals) legislators, managers, service workers, market workers and army man.

The change of expensive cigarette prices positively and significantly influences a log of smoked tobacco that confirms conclusions in sections 4 and 5. In model specification (Table A12) with the Goskomstat price variable — *pricegks* negative price elasticity of consumed tobacco amount proves to be true.

Finally, we came to the conclusion that in 2000–2001 people began to smoke on the average, 10% more.

7. COMPOSITE MODEL WITH DEPENDENCE BETWEEN QUITTING AND CONSUMPTION

More complicated tools of econometric modelling for construction of interrelated models with dependence between consumption and quitting of cigarettes in view of overcoming the first hurdle — the beginnings of smoking are used.

More complicated method of accounting in smoking termination model of unobserved heterogeneity consists in applying Cox–Snell cumulative residuals as a covariate got from the smoking initiation model. We can take into account double-hurdle sample selectivity in consumption model: the first hurdle — beginning of smoking, the second hurdle — quitting of smoking. For this purpose we shall include into consumption model the predicted hazards, as the explanatory variables, received in models of initiation and quitting of smoking. For correction on unobserved heterogeneity in starting and quitting we use Cox–Snell residuals from models of initiation and quitting of smoking. Besides, to model consumption habit, we use lagged fitted consumption as the explaining variable.

Thus, the composite model with dependence between quitting and consumption has been constructed proceeding from exogeneity of smoking initiation and endogeneity of smoking consumption and quitting according to the following algorithm.

1. We estimate Cox's model of smoking initiation and receive Cox–Snell residuals res_{start} and hazard prediction $hazard_{start}$.

2. We construct the model of smoked cigarettes amount, using res_{start} and $hazard_{start}$ as explaining variables. We receive the fitted value[10] of smoked cigarettes amount $\hat{q}_{it}^{k}$ and residuals of model res_{cons}.

3. We estimate Cox's model for smoking termination, using res_{start}, res_{cons} and $\hat{q}_{it}^{k}$ as covariates. We receive Cox–Snell residuals res_{stop} and hazard prediction $hazard_{stop}$.

4. We construct the model of smoked cigarettes amount, using res_{start}, res_{stop}, $hazard_{start}$, $hazard_{stop}$ and lagged fitted consumption $\hat{q}_{it-1}^{k}$ as explaining variables. Again we receive the fitted value of

[10] Here k — iteration number.

amount of consumed cigarettes $\hat{q}_{it-1}^{k+1}$ and residuals of model res_{cons}. We go to step 3, using again the received estimates.

5. We repeat steps 3 and 4.

Construction of model of smoked cigarettes amount with Cox–Snell residuals and predicted hazards as explaining variables is being carried out on the basis of panel data also.

The results received for these composed models of smoking termination and consumption are presented in Tables 5 and 6 accordingly (after 3 iterations).

The graphic of cumulative Cox–Snell residuals for the model submitted in Table 5, shows satisfactory fitting. The predicted value of quitted to smoke makes 14% at observed in sample of 21%.

Sample consists only of a part of those individuals (there is a complete set of explaining variables), who started smoking in the period from 1993 till 2001 or quitted smoking at the same period or continued to smoke. Some variables are not included into the model of smoking termination (Table 5), for example, *selfhealth*, to increase the sample volume.

Table 5. Results of Cox's estimates for composite model: quitted smoking.

Variables	Coefficients
Female	–0.001 (0.321)
Marst	0.146 (0.157)
Durationcat	-0.444^{***} (0.048)
Sport	–0.082 (0.065)
Pricel	1.546^{***} (0.592)
Priceu	0.250 (0.308)
$\hat{q}$	-1.611^{***} (0.437)
res_{start}	-0.778^{**} (0.407)
res_{cons}	–6.684 (5.390)
No. of subjects	2573
No. of failures	543
Log likelihood	–3081.01
χ^2	1588.40
Predicted proportion of stopped	0.14
Observed proportion of stopped	0.21

The values of coefficients are presented in the table. Breslow method for ties. The robust standard errors adjusted for clustering on individuals are in brackets. Significance of coefficients: * — 10% level, ** — 5% level, *** — 1% level.

Table 6. Results of composite tobit model: logarithm of amount of smoked cigarettes (RLMS, 1994–2001).

Variables	Coefficients
Female	-0.538^{***} (0.161)
Marst	–0.051 (0.052)
Age	-0.026^{**} (0.013)
Agesq	0.00001 (0.0002)
Educat	0.015 (0.015)
Occat	0.002 (0.007)
Sort	0.404^{***} (0.022)
Pricel	–0.026 (0.114)
Priceu	–0.023 (0.085)
Sport	-0.023^{*} (0.013)
Work	0.111^{***} (0.046)
res_{start}	1.773^{***} (0.094)
$hazard_{start}$	0.926^{***} (0.185)
res_{stop}	-0.349^{**} (0.161)
$hazard_{stop}$	$3.98\cdot10^{-8}$ ($4.23\cdot10^{-8}$)
$\hat{q}_{t-1}$	0.040^{***} (0.004)
Log likelihood	–2046.94
Wald χ^2	2982.63
Likelihood-ratio test of sigma_u=0, χ^2	320.68
sigma_u	0.538
sigma_e	0.428
rho	0.612
Observation:	
uncensored	2018
left-censored	555

The standard errors are in brackets. Significance of coefficients: * — 10% level, ** — 5% level, *** — 1% level. Occat — categorical variable of profession of individual.

The amount of significant determinants was noticeably reduced as compared to that one, presented in Table 3. Cheap cigarettes price variable is positive and expensive cigarettes — insignificant. The residuals of consumption model are insignificant. We received the expected negative elasticity of hazard smoking termination by initiation model residuals and consumption estimation $\hat{q}_{ti}^{k}$, that confirms the interrelation between amount of consumed cigarettes and decision to quit smoking.

A part of in the model of smoked cigarettes amount are significant with expected signs: sex, age, physical activity and type of tobacco. However, the price variables are insignificant. The expected positive sign is received in significant coefficient at lagged estimation of tobacco consumption $\hat{q}_{t-1}$ that allows to make a conclusion about addictive character of cigarettes consumption, *viz.* the increase of smoked cigarettes amount by 1 pc nowadays, results in growth of consumption by 0.04 pcs tomorrow. Residuals of smoking initiation and termination models is significant with expected signs and are interpreted as a increase/reduction of amount of smoked cigarettes when the number of starting/quitting smokers is growing. The factor of the first hurdle — risk of the beginning of smoking is significant. Hence censoring individuals do not smoke because they have never smoked.

8. CONCLUSION

The problems of behavior of tobacco customers in Russia are insufficiently studied even at the empirical level. We have made an attempt to analyse the reasons determining the overcoming of a double-hurdle by individuals, *viz.* smoking initiation and termination. We have received the following basic results.

- Social and economic changes have led to a wide spreading of a smoking habit at the beginning of the current century. The number of smokers has grown 3.6% from 1994 till 2001 and 2% in 2000–2001; moreover, the amount of smoked cigarettes has increased average 10%.
- Approximately 35% of smokers stop smoking per the first 5 years after the beginning of smoking. The estimations have proved that among quitted-smoking teenagers about 90% quit to smoke per the first 3 years after the beginning of smoking.
- Price is the key factor for the beginning and termination of smoking. Licensing of sellers, planned by the Government, will inevitably "wash away" small trading firms and lead to increase of prices for cigarettes. The asymmetric influence on price increase of cigarettes according to their sorts has been revealed. The increase of price on cheap cigarettes results in raise of probability of smoking beginning of available cigarettes by non-smokers due to the falling of living standards and simultaneously it leads more often to smoking failure due to low incomes of smokers. Expensive cigarettes price increase reduces the probability of smoking beginning for non-smokers and lowers the frequency of smoking quitting for smokers; cigarettes price is minor to rich individuals as compared with their habit of smoking; current cigarettes price, probably, impacts not so much the termination of smoking, but mostly the shift in demand for

better quality of cigarettes. Thus, to regulate tobacco market it is necessary to take into account the heterogeneity of price elasticity of tobacco consumption.

- The cheap cigarettes prices increase results in growth of probability of smoking beginning with the youth during a short-time period almost by 0.03 that testifies the presence of a social problem — declining of life standards together with the growth of real prices leads to wide practise of a harmful habit of smoking among teenagers.
- The differences in reasons for smoking beginning with women and men have been revealed. Termination of smoking for women is more characteristic during the fertility periods; women have the greater elasticity of factors connected with self-estimation of health and quantitative family structure. The effect of generation is important for men.
- Townspeople are more inclined to smoking beginning since public opinion in countryside still remains the influential factor and hinders smoking because smoking people, especially women, are considered violaters of social norms. At the same time, public opinion in cities concerning harmful habits is more democratic.
- Empirically addictive character of cigarettes consumption proves to be true: the increase of smoked cigarettes amount by 1 pc today results in growth of consumption by 0.04 pcs tomorrow. Experienced smokers are less inclined into smoking. The elasticity of risk of smoking failure for women and men, according to smoking duration, is approximately equal.
- Aging is a good reason for reduction of cigarettes consumption: the domed quadratic dependence of smoked cigarettes quantity upon age of the individual has been obtained with the extremum point at the age about 29, after which the quantity of smoked cigarettes begins to reduce.
- Obtained results of our research allow to judge on a positive effect of restrictive legislative measures and, in particularly, on a positive influence of prohibition to smoke in public places (at worksites) and anti-smoking propaganda. Intensive work and employment essentially influence the quantity of consumed tobacco.
- The possibilities of smoking reduction have been found through propagation of a healthy lifestyle since physical activity restrains the use of tobacco.

APPENDIX

Table A1. Economics of tobacco — country data report by World Bank database.

Topics	Russia	Ukraine	USA	Germany	Switzerland	China	France	Spain
Smoking Prevalence, Men (1993), %	67	56.5	27.7	36.8	36	61	40	48
Smoking Prevalence, Women (1993), %	30	21.9	25.5	21.5	26	7	27	25
Cigarette Consumption Per Capita (in pieces, 1995)	1756.85	1937.33	2296.2	2174.96	2674.58	1864.58	1885.31	2384.42
Mortality Percent From Smoking (Ages 35–69, All Cancer,1995), %	40	39	42	31	29	NA	34	34
Mortality Percent From Smoking (Ages 35–69,Vascular Disease,1995), %	32	29	33	22	20	NA	23	21

Source: http://www1.worldbank.org/tobacco/database.asp.

Table A2. Variable definition and sample means.

Variables	Definition	Starting		Quitting	
		Men	Women	Men	Women
Marst	=1 if married	0.42	0.54	0.67	0.52
Urban	=1 if urban of place of residence	0.69	0.68	–	–
Sport	Variants of physical activities, 1 — Light physical exercise for relaxation, less than three times a week, 2 — Medium and intensive physical exercise, less than three times a week, 3 — Intensive physical exercise at least three times a week 15 minutes or more, 4 — Daily exercise not less than 30 minutes a day, 0 — Does not engage in physical activities	–	–	0.58	0.52
Numhh	Number of household members	3.71	3.36	–	–
Selfhealth	Lagged health self-estimation: 1 — very good, 2 — good, 3 — average, not good, but not bad, 4 — bad, 5 — very bad	2.43	2.87	2.73	2.86
Prof	Professional categories, =1 if legislators, senior managers, officials, professionals, clerks, technicians and associate professsionals	–	–	0.23	0.46
Durationcat	Categories of smoking duration, 1 — <5 years, 2 — 6–10 years, 3 — 11–20 years, 4 — 21–30 years, 5 — 31–40 years, 6 — 41–50 years, 7 — 51+ years	–	–	3.15	2.12

Variables	Definition	Starting		Quitting	
		Men	Women	Men	Women
Pricegks	Prices for 20 cigarettes pack	0.43	0.49	0.26	0.29
Pricel	Prices for 20 cheap cigarettes pack	0.26	0.25	0.28	0.28
Priceu	Prices for 20 expensive cigarettes pack	0.39	0.39	0.38	0.42
Age start		–	–	16.48	19.66
Average smoking duration for former smokers, years		–	–	15.17	5.82
Average quantities of cigarettes per day for smokers		–	–	15.18	8.62
Number of observations		1129	3669	1492	434
Number of failures		250	149	82	56

Table A3. Smokers by year, RLMS.

	1994	1995	1996	1998	2000	2001
Men	0.82	0.83	0.81	0.80	0.78	0.75
Women	0.18	0.17	0.19	0.20	0.22	0.25
Total	1	1	1	1	1	1
Smokers under age 30						
Men	0.74	0.76	0.75	0.74	0.73	0.70
Women	0.26	0.24	0.25	0.26	0.27	0.30
Total	1	1	1	1	1	1

Table A4. Smoking duration for smokers by age (RLMS, 1994–2001).

Age categories	Average duration, years
14–20	2.7
21–26	6.4
27–30	10.4
31–40	16.2
41–50	24.1
51–65	35.4

Table A5. Real cigarette price data, rbl. per 1 pack (Goskomstat, 1993–2001).

# Primary sample unit	1993	1994	1995	1996	1997	1998	1999	2000	2001
1	0.04	0.09	0.24	0.26	0.27	0.40	0.36	0.34	0.58
2	0.05	0.09	0.25	0.22	0.22	0.27	0.24	0.22	0.44
3	0.07	0.14	0.38	0.34	0.34	0.46	0.43	0.38	0.65
4	0.06	0.11	0.30	0.28	0.28	0.38	0.31	0.31	0.51
5	0.06	0.11	0.30	0.28	0.28	0.38	0.31	0.31	0.51
6	0.04	0.08	0.23	0.27	0.27	0.32	0.28	0.26	0.44
7	0.08	0.15	0.42	0.34	0.35	0.50	0.44	0.41	0.75
8	0.06	0.11	0.31	0.28	0.32	0.46	0.39	0.36	0.76
9	0.07	0.14	0.39	0.38	0.44	0.83	0.54	0.45	0.73
10	0.08	0.15	0.41	0.34	0.34	0.43	0.46	0.31	0.59
11	0.08	0.16	0.44	0.44	0.41	0.52	0.43	0.35	0.69
12	0.07	0.14	0.39	0.33	0.38	0.62	0.51	0.49	0.82
13	0.06	0.11	0.32	0.31	0.33	0.39	0.32	0.30	0.54
14	0.07	0.13	0.37	0.32	0.35	0.52	0.41	0.34	0.53
15	0.07	0.13	0.37	0.34	0.40	0.47	0.26	0.33	0.60
16	0.04	0.08	0.21	0.22	0.30	0.49	0.45	0.39	0.58
17	0.04	0.08	0.23	0.21	0.28	0.52	0.46	0.38	0.66
18	0.04	0.08	0.23	0.21	0.28	0.52	0.46	0.38	0.66
19	0.06	0.12	0.33	0.31	0.31	0.49	0.42	0.36	0.64
20	0.10	0.18	0.51	0.47	0.46	0.73	0.63	0.56	0.85
21	0.07	0.13	0.36	0.30	0.35	0.49	0.46	0.43	0.67
22	0.06	0.11	0.32	0.31	0.39	0.61	0.52	0.48	0.69
23	0.08	0.15	0.42	0.42	0.44	0.63	0.54	0.51	0.77
24	0.06	0.11	0.32	0.31	0.39	0.61	0.52	0.48	0.69
25	0.14	0.27	0.75	0.31	0.33	0.46	0.44	0.34	0.65
26	0.06	0.12	0.33	0.35	0.34	0.48	0.41	0.34	0.53
27	0.05	0.09	0.27	0.26	0.33	0.59	0.41	0.38	0.57
28	0.06	0.11	0.31	0.41	0.41	0.55	0.50	0.39	0.64
29	0.07	0.13	0.38	0.34	0.30	0.46	0.37	0.32	0.59
30	0.14	0.27	0.75	0.31	0.33	0.46	0.44	0.34	0.65
31	0.06	0.11	0.30	0.28	0.31	0.54	0.51	0.45	0.74
32	0.02	0.04	0.11	0.16	0.19	0.42	0.33	0.30	0.44
33	0.08	0.14	0.40	0.38	0.45	0.57	0.61	0.48	0.88
34	0.08	0.14	0.40	0.38	0.45	0.57	0.61	0.48	0.88
35	0.05	0.09	0.24	0.22	0.26	0.33	0.29	0.28	0.48
36	0.05	0.10	0.29	0.27	0.34	0.38	0.32	0.30	0.52
37	0.05	0.09	0.24	0.22	0.26	0.33	0.29	0.28	0.48
38	0.06	0.11	0.30	0.31	0.47	0.50	0.49	0.44	0.73

Table A6. Log-rank test for equality of survivor functions (χ^2).

Variables	Initiation smoking	Stopped smoking
Female	388.18	12.88
Marst	160.90	0.22
Urban	2.84	–
Numhh	51.40	–
Prof	–	17.92
Selfhealth	69.12	3.34
Sport	–	10.63
Durationcat	–	73.38
Pricel	514.60	119.83
Priceu	784.72	185.71
Pricegks	1388.89	239.00

Table A7. Cox's model results for smoking initiation (Goskomstat price).

Variables	Full sample	Men	Women
Female	-0.426^{***} (0.052)	–	–
Marst	-0.099^{*} (0.058)	0.027 (0.081)	–0.117 (0.082)
Urban	0.355^{***} (0.057)	0.469^{***} (0.074)	0.233^{**} (0.093)
Pricegks	-11.94^{***} (0.255)	-10.366^{***} (0.320)	-13.876^{***} (0.395)
Numhh	–0.018 (0.018)	0.023 (0.023)	-0.062^{**} (0.025)
30<Age≤50	-1.339^{***} (0.118)	-1.434^{***} (0.204)	-1.232^{***} (0.145)
Age >50	-3.365^{***} (0.328)	-3.236^{***} (0.574)	-3.264^{***} (0.384)
1995	1.083^{***} (0.295)	1.047^{**} (0.422)	1.134^{***} (0.412)
1996	1.113^{***} (0.274)	0.978^{**} (0.400)	1.254^{***} (0.384)
1998	2.192^{***} (0.245)	1.845^{***} (0.365)	2.506^{***} (0.330)
2000	1.475^{***} (0.254)	1.388^{***} (0.376)	1.428^{***} (0.354)
2001	2.624^{***} (0.235)	2.125^{***} (0.355)	3.144^{***} (0.315)
No. of subjects	8445	2254	6191
No. of failures	1391	792	599
Log likelihood	–9627.53	–4882.04	–3746.95
χ^2	3528.89	1359.39	2085.04
Predicted proportion of starters	0.04	0.09	0.02
Observed proportion of starters	0.10	0.22	0.06

The values of coefficients are presented in the table. Breslow method for ties. The robust standard errors adjusted for clustering on individuals are in brackets. Significance of coefficients: * — 10% level, ** — 5% level, *** — 1% level.

Table A8. Test of proportional hazards assumption (Pr>χ^2).

Variables	Full sample	Men	Women
Female	0.467	–	–
Selfhealth	0.518	0.883	0.416
Marst	0.002	0.002	0.216
Urban	0.291	0.949	0.112
Pricel	0.878	0.511	0.287
Priceu	0.067	0.214	0.066
Numhh	0.576	0.337	0.970
30<Age≤50	0.400	0.300	0.561
Age >50	0.306	0.096	0.729
1995	0.606	0.768	0.758
1996	0.488	0.506	0.786
1998	0.319	0.417	0.377
2000	0.186	0.332	0.341
2001	0.484	0.525	0.564
Global test	0.070	0.092	0.721

The rank of analysis time is used.

Table A9. Test of proportional hazards assumption under age 20 (Pr>χ^2).

Variables	Full sample	Men	Women
Female	0.452	–	–
Selfhealth	0.383	0.524	0.395
Marst	0.322	0.678	0.287
Urban	0.542	0.525	0.429
Pricel	0.664	0.548	0.861
Priceu	0.047	0.129	0.229
1994	0.795	0.842	0.820
1995	0.863	0.705	0.955
1996	0.462	0.675	0.238
2000	0.286	0.368	0.551
2001	0.991	0.915	0.826
Global test	0.128	0.403	0.651

The rank of analysis time is used.

Table A10. Cox's model results for quitted smoking (Goskomstat price).

Variables	Full sample	Men	Women
Female	0.229 (0.163)	–	–
Selfhealth	0.030 (0.125)	0.148 (0.156)	–0.174 (0.211)
Marst	0.191 (0.152)	0.121 (0.208)	0.194 (0.229)
Sport	0.064 (0.063)	0.155** (0.072)	–0.189 (0.138)
Durationcat	–0.517*** (0.088)	–0.515*** (0.121)	–0.534*** (0.135)
Pricegks	–11.023*** (0.818)	–10.228*** (1.137)	–12.263*** (1.282)
Prof	0.512*** (0.154)	0.711*** (0.208)	0.256 (0.224)
Lnincome	–0.034* (0.017)	–0.021 (0.022)	–0.053* (0.030)
30<Age≤50	0.112 (0.196)	0.203 (0.279)	0.027 (0.289)
Age >50	0.729* (0.381)	0.877* (0.515)	0.541 (0.667)
1994	–0.490 (0.436)	–0.676 (0.518)	0.032 (0.876)
1995	0.550 (0.496)	–0.118 (0.606)	1.636* (0.954)
1998	0.975* (0.408)	0.815* (0.497)	1.362* (0.791)
2000	–0.184 (0.402)	–0.171 (0.481)	–0.040 (0.791)
2001	0.290 (0.398)	0.001 (0.501)	0.737 (0.766)
No. of subjects	1996	1529	467
No. of failures	208	119	89
Log likelihood	–1273.13	–712.36	–413.57
χ^2	366.19	180.16	181.63
Predicted proportion of stopped	0.02	0.01	0.01
Observed proportion of stopped	0.10	0.08	0.19

The values of coefficients are presented in the table. Breslow method for ties. The robust standard errors adjusted for clustering on individuals are in brackets. Significance of coefficients: * — 10% level, ** — 5% level, *** — 1% level.

Table A11. Test of proportional hazards assumption (Pr>χ^2).

Variables	Full sample	Men	Women
Marst	0.095	0.004	0.663
Selfhealth	0.256	0.266	0.295
Female	0.094	–	–
Sport	0.949	0.918	0.325
Durationcat	0.656	0.127	0.166
Pricel	0.078	0.039	0.742
Priceu	0.407	0.179	0.389
Prof	0.803	0.252	0.286
Lnincome	0.314	0.640	0.605
30<Age≤50	0.156	0.466	0.086
Age >50	0.116	0.609	0.317
1995	0.904	0.322	0.889
1996	0.670	0.251	0.321
1998	0.612	0.473	0.551
2000	0.216	0.125	0.402
2001	0.807	0.684	0.802
Global test	0.253	0.099	0.763

The rank of analysis time is used.

Table A12. Tobit random-effects regression for logarithm of amount of smoked cigarettes (Goskomstat price).

Variables	RLMS panel 1995–2001		RLMS panel 1994–2001	
	Full sample	For employment	Full sample	For employment
Const	–0.609*** (0.118)	–0.648*** (0.123)	–0.926*** (0.103)	–0.894*** (0.106)
Marst	–0.068*** (0.021)	–0.063*** (0.022)	–0.064*** (0.019)	–0.054*** (0.021)
Female	–0.879*** (0.027)	–0.850*** (0.028)	–0.910*** (0.026)	–0.879*** (0.027)
Age	0.016*** (0.006)	0.014*** (0.006)	0.020*** (0.005)	0.017*** (0.006)
Agesq	–0.0003*** (0.00007)	–0.0002*** (0.00008)	–0.0003*** (0.00007)	–0.0003*** (0.00007)
Elementary Education	–0.032 (0.031)	–0.039 (0.032)	–0.012 (0.028)	–0.016 (0.028)
PTU with Secondary Education	0.032 (0.024)	0.040 (0.026)	0.033 (0.022)	0.034 (0.023)
Vocational Education	–0.007 (0.028)	–0.015 (0.028)	–0.009 (0.026)	–0.015 (0.027)

Variables	RLMS panel 1995–2001		RLMS panel 1994–2001	
	Full sample	For employment	Full sample	For employment
Higher Education	–0.067** (0.034)	–0.075** (0.035)	–0.056* (0.031)	–0.064** (0.032)
Legislators, Senior Managers, Officials	0.161*** (0.043)	0.202*** (0.045)	0.156*** (0.040)	0.195*** (0.042)
Professionals	–0.046 (0.039)	–0.025 (0.040)	–0.034 (0.035)	–0.016 (0.037)
Clerks	0.013 (0.052)	0.009 (0.054)	0.029 (0.048)	0.020 (0.050)
Service Workers&Market Workers	0.182*** (0.040)	0.140*** (0.042)	0.196*** (0.037)	0.176*** (0.039)
Craft&Related Trades	–0.007 (0.035)	–0.008 (0.037)	0.004 (0.033)	0.008 (0.034)
Plant&Machine Operators&Assemblers	0.033 (0.035)	0.009 (0.037)	0.049 (0.033)	0.030 (0.034)
Unskilled	0.008 (0.038)	–0.020 (0.039)	0.039 (0.035)	0.021 (0.037)
Army	0.226*** (0.078)	0.231*** (0.083)	0.203*** (0.071)	0.184** (0.076)
Sort	1.339*** (0.010)	1.350*** (0.011)	1.310*** (0.009)	1.325*** (0.009)
Pricegks	–0.738*** (0.084)	–0.742*** (0.087)	–0.723*** (0.083)	–0.733*** (0.086)
Sport	–0.018** (0.008)	–0.022*** (0.008)	–	–
Hours	–	0.014*** (0.002)	–	0.013*** (0.002)
Empl	0.050* (0.028)	–	0.065*** (0.025)	–
1995	–0.027 (0.025)	–0.018 (0.027)	0.222*** (0.027)	0.164*** (0.030)
1996	–0.025 (0.026)	–0.019 (0.028)	0.234*** (0.026)	0.177*** (0.029)
1998	–	–	0.255*** (0.037)	0.192***(0.039)
2000	0.088*** (0.023)	0.076*** (0.025)	0.349*** (0.029)	0.274*** (0.033)
2001	0.377*** (0.024)	0.374*** (0.026)	0.631*** (0.047)	0.567*** (0.049)
Log likelihood	–12205.06	–10992.25	–14849.39	–13346.29
Wald χ^2	23062.28	21337.01	26765.71	24813.11
Likelihood-ratio test of sigma_u=0, χ^2	2275.33	1920.51	2983.89	2493.32
sigma_u	0.621	0.610	0.625	0.613
sigma_e	0.618	0.618	0.629	0.629
rho	0.502	0.494	0.497	0.487
Observation:				
uncensored	8484	7652	10277	9236
left-censored	13518	12006	16677	14772

Standard errors are in brackets. Significance of coefficients: * — 10% level, ** — 5% level, *** — 1% level. Base categories: Secondary Education, Technicians and Associate Professionals.

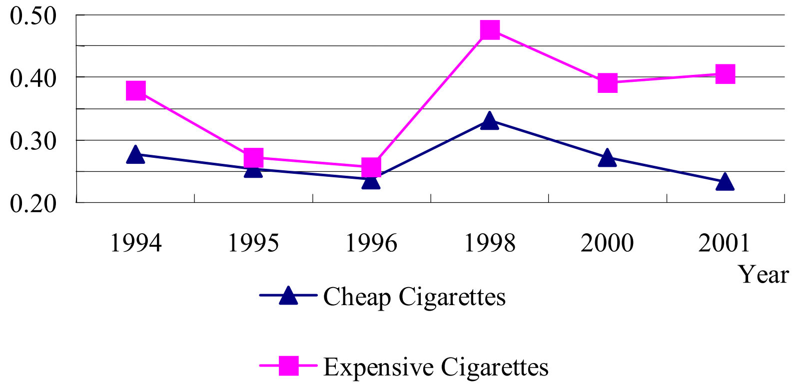

Fig. A1. Deflated cigarette price, rbl. per 1 pack (RLMS).

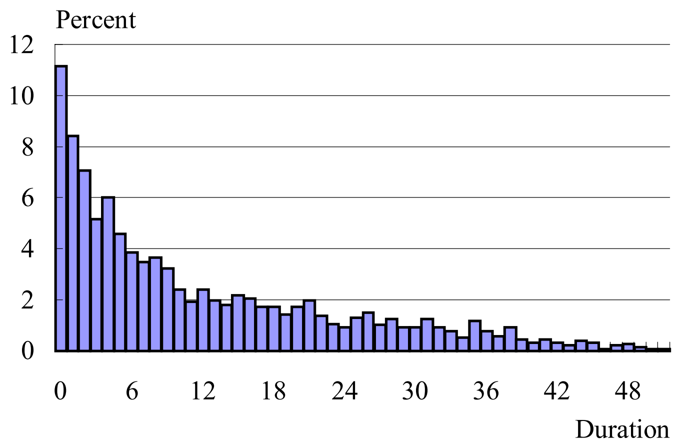

Fig. A2. Distribution of smoking duration for stopped smokers, RLMS.

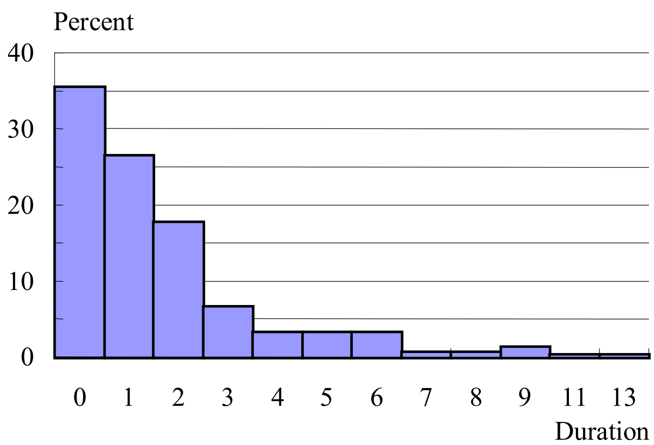

Fig. A3. Distribution of smoking duration for stopped smokers (under age 20), RLMS.

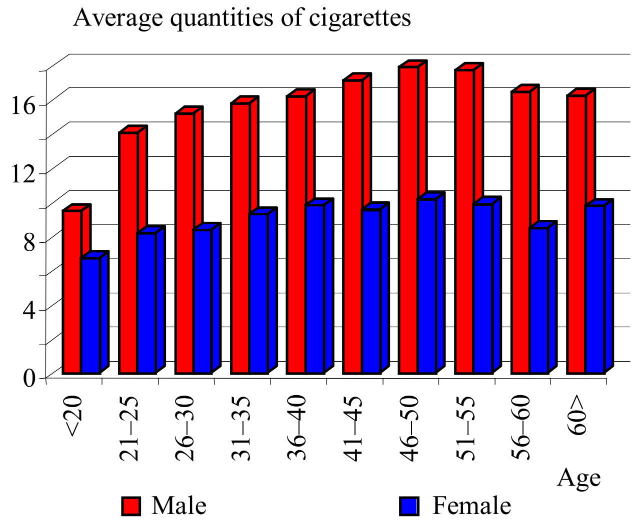

Fig. A4. Average quantities of daily smoked cigarettes.

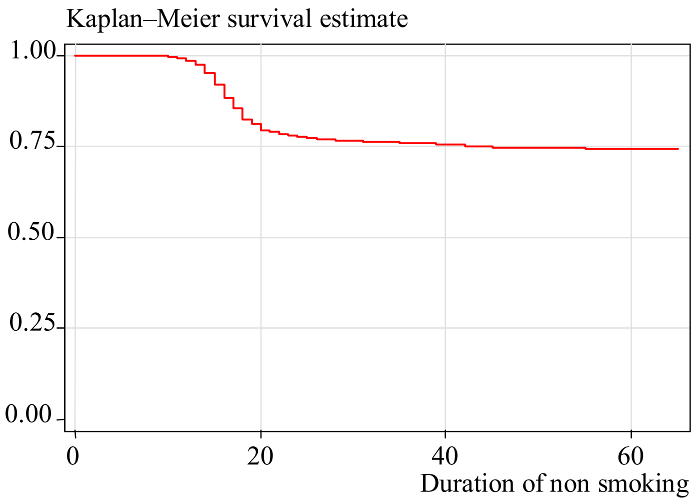

Fig. A5. Estimate of the survival function for starting smoking (duration of non smoking), full sample.

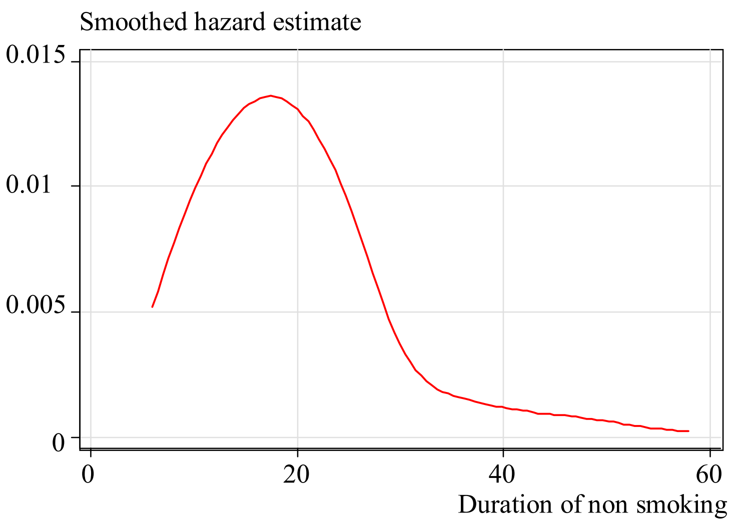

Fig. A6. Estimate of the hazard function for starting smoking (duration of non smoking), full sample.

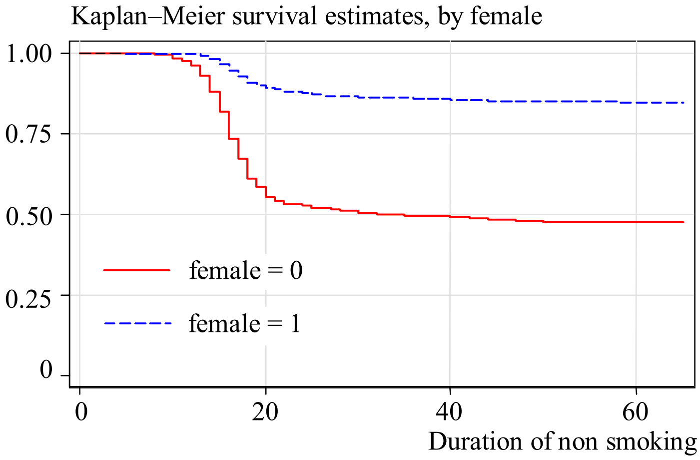

Fig. A7. Estimate of the survival function for starting smoking (duration of non smoking) by gender.

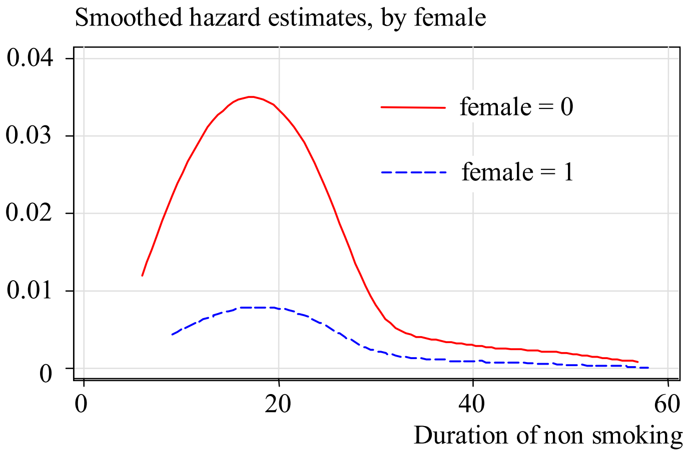

Fig. A8. Estimate of the hazard function for starting smoking (duration of non smoking) by gender.

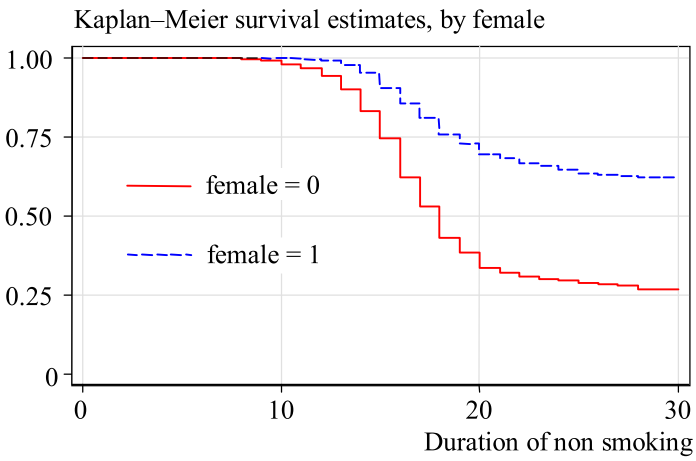

Fig. A9. Estimate of the survival function for starting smoking by gender, under age 30.

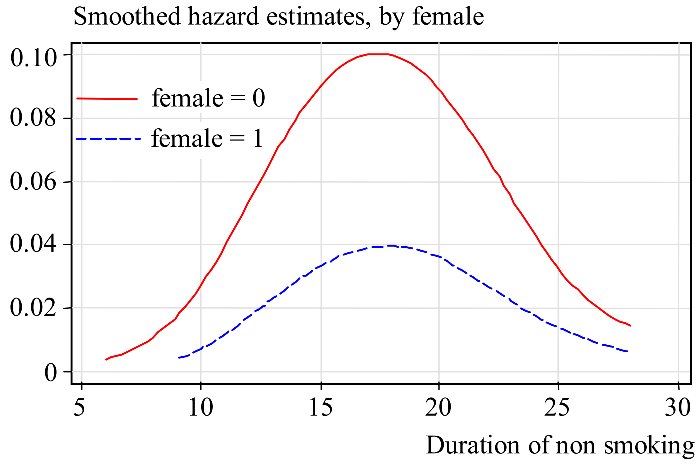

Fig. A10. Estimate of the hazard function for starting smoking by gender, under age 30.

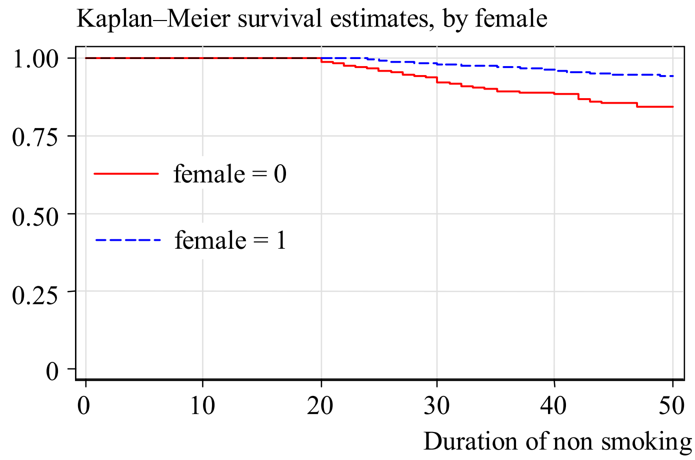

Fig. A11. Estimate of the survival function for starting smoking by gender, age 31–50.

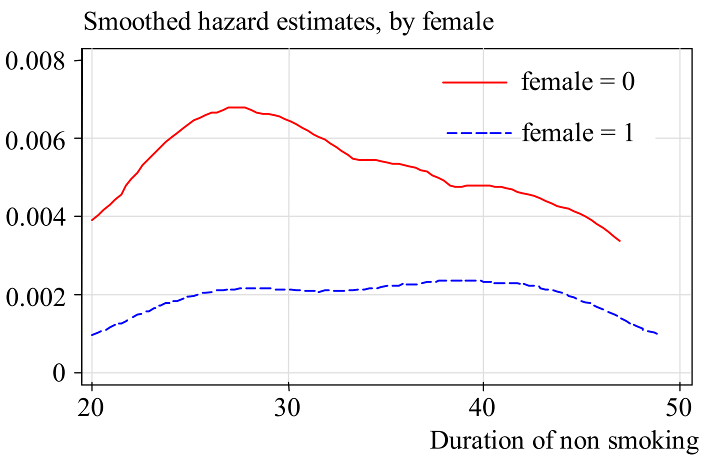

Fig. A12. Estimate of the hazard function for starting smoking by gender, age 31–50.

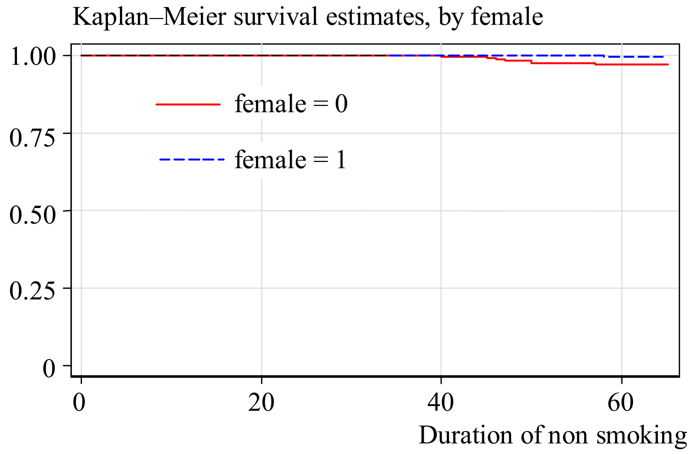

Fig. A13. Estimate of the survival function for starting smoking by gender, after age 50.

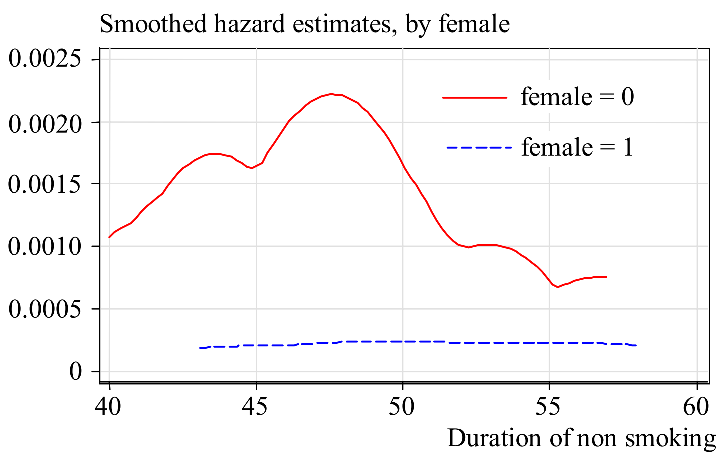

Fig. A14. Estimate of the hazard function for starting smoking by gender, after age 50.

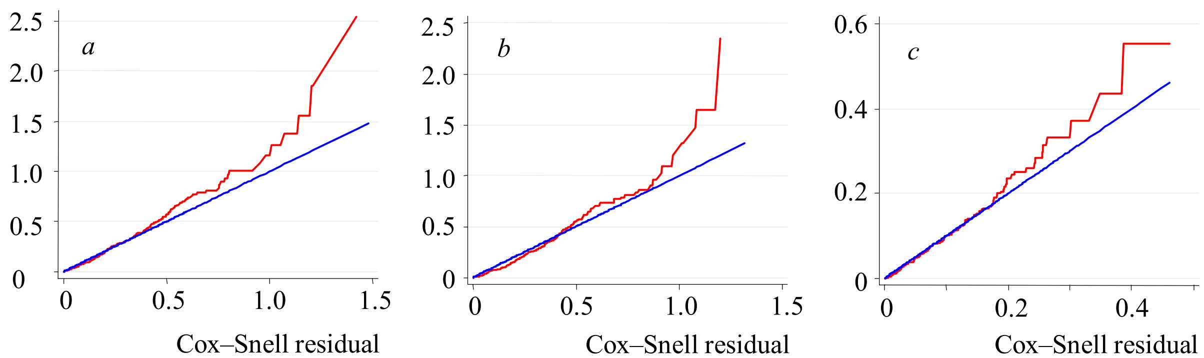

Fig. A15. Cumulative Cox–Snell residuals for starting smoking, base model: full sample (*a*), men (*b*), women (*c*).

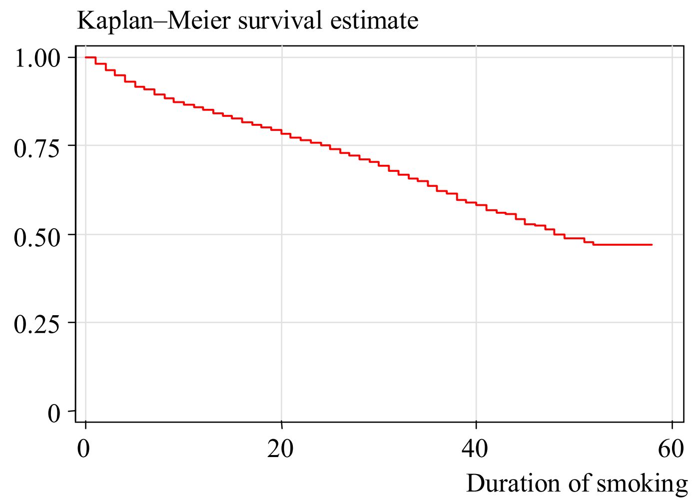

Fig. A16. Estimate of the survival function for stopped smoking (duration of smoking), full sample.

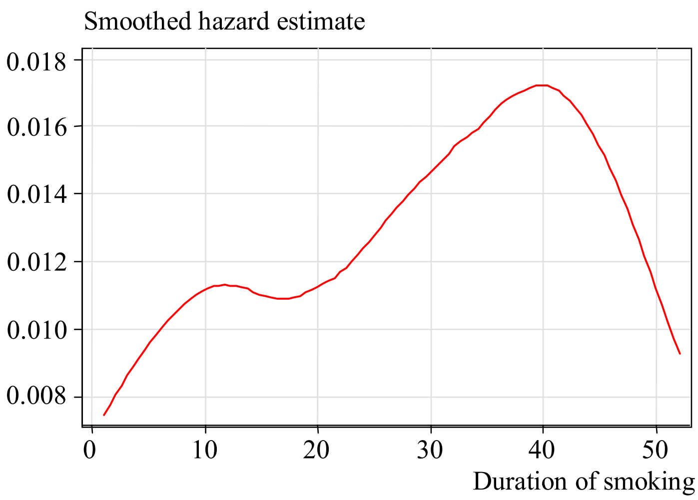

Fig. A17. Estimate of the hazard function for stopped smoking (duration of smoking), full sample.

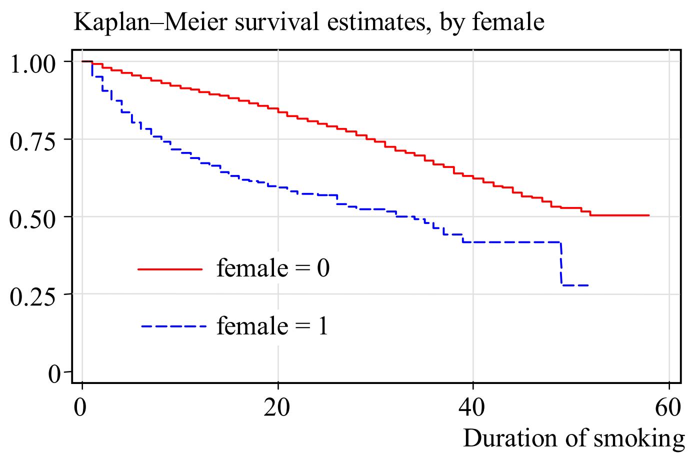

Fig. A18. Estimate of the survival function for stopped smoking (duration of smoking), by gender.

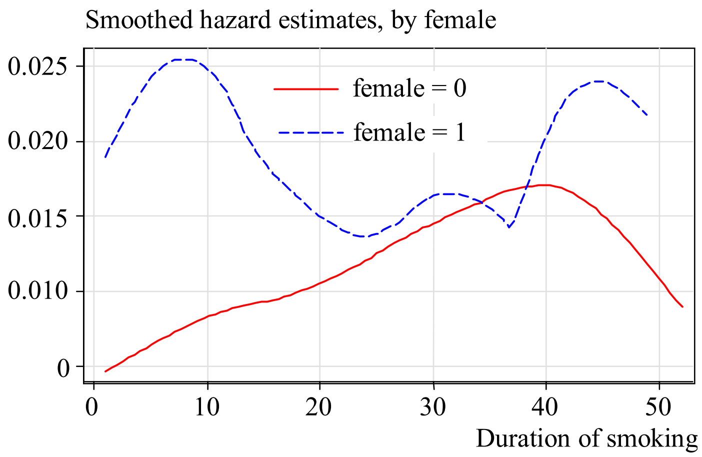

Fig. A19. Estimate of the hazard function for stopped smoking (duration of smoking), by gender.

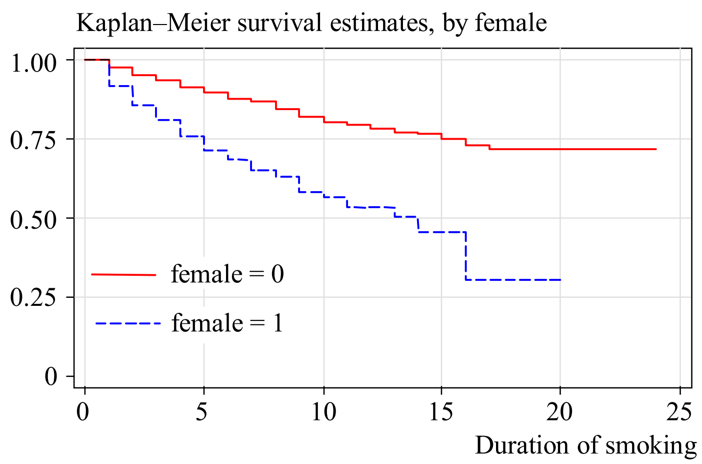

Fig. A20. Estimate of the survival function for stopped smoking by gender, under age 30.

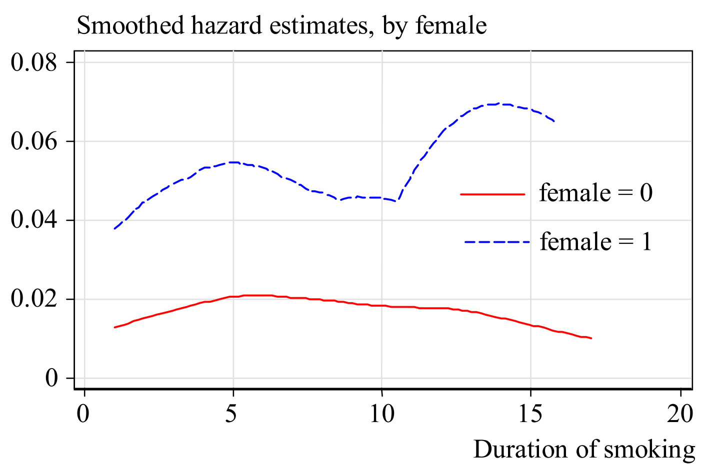

Fig. A21. Estimate of the hazard function for stopped smoking by gender, under age 30.

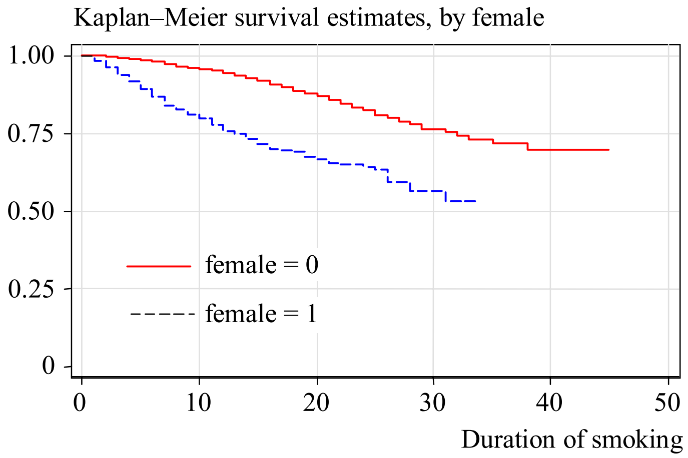

Fig. A22. Estimate of the survival function for stopped smoking by gender, age 31–50.

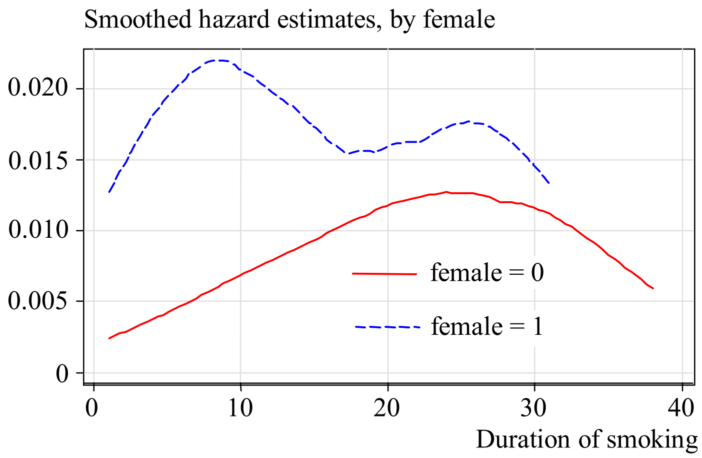

Fig. A23. Estimate of the hazard function for stopped smoking by gender, age 31–50.

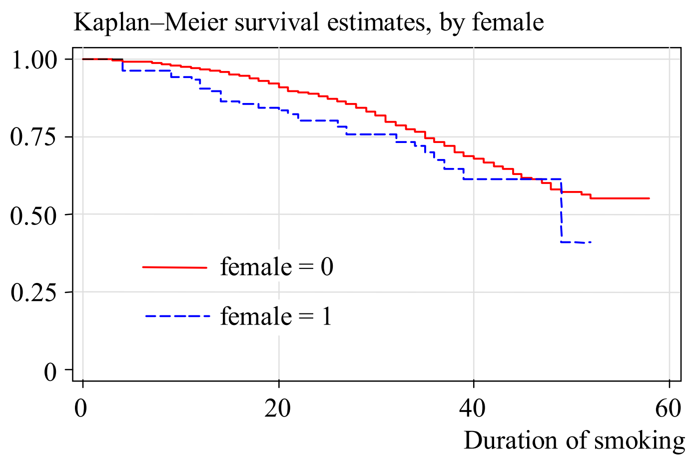

Fig. A24. Estimate of the survival function for stopped smoking by gender, after age 50.

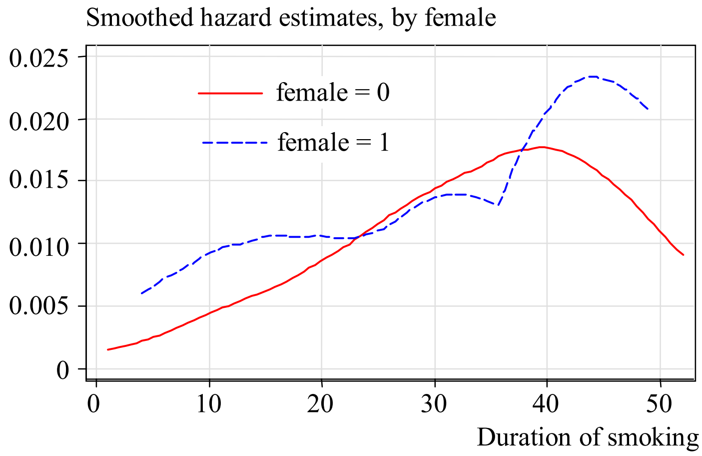

Fig. A25. Estimate of the hazard function for stopped smoking by gender, after age 50.

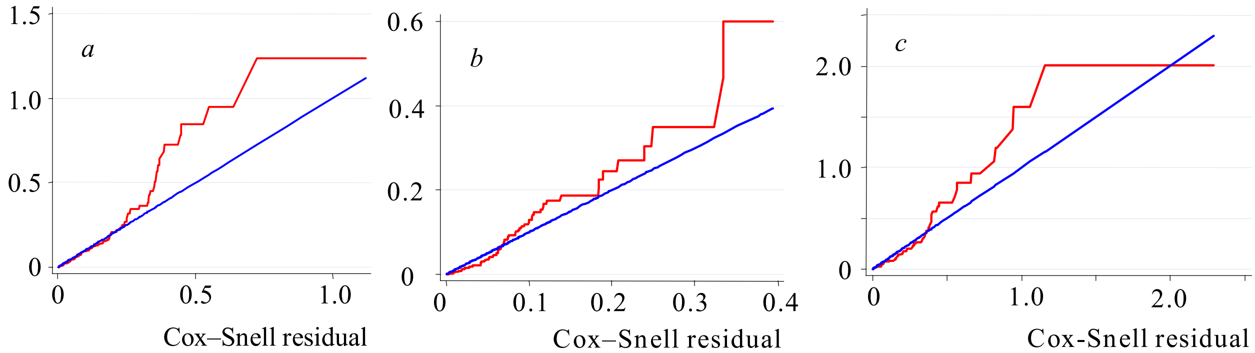

Fig. A26. Cumulative Cox–Snell residuals for stopped smoking, base model: full sample (*a*), men (*b*), women (*c*).

REFERENCES

Becker, G.S. and K.M. Murphy (1988) A theory of rational addiction, *Journal of Political Economy* **96** (4), 675–700.

Becker G.S., M. Grossman, and K.M. Murphy (1994) An Empirical Analysis of Cigarette Addiction, *American Economic Review* **84** (3), 396–418.

Beenstock, M. and G. Rahav (2002) Testing Gateway Theory: The Effect of Cigarette Prices on the Consumption of Cigarettes and Illicit Drugs in Israel, *Journal of Health Economics* **21**, 679–698.

Beenstock, M. and G. Rahav (2003) *Immunity and Susceptibility in Illicit Drug Initiation in Israel*, Unpublished paper.

Chaloupka, F.J. and K.E. Warner (1999) The Economics of Smoking, *Handbook of Health Economics* **1B** (Elsevier), 1539–1627.

Cox, D.R. (1972) Regression Models and Life-Tables, *Journal of the Royal Statistical Society*, Series B, **34** (2), 187–202.

Cox, D. and D. Oakes (1984) *Analysis of Survival Data* (London: Chapman & Hall).

Douglas, S.M. (1998) The Duration of the Smoking Habit, *Economic Inquiry* **36** (1), 49–64.

Evans, W.N. and M.C. Farrelly (1998) The compensating behavior of smokers: taxes, tar and nicotine. *RAND Journal of Economics* **29** (3), 578–595.

Evans, W.N., M.C. Farrelly and E. Montgomery (1999) Do Workplace Smoking Bans Reduce Smoking? *American Economic Review* **89** (5), 729–747.

Forster, M., and A.M. Jones. (2001) The Role of Tobacco Taxes in Starting and Quitting Smoking: Duration Analysis of British Data. *Journal of the Royal Statistical Society*, Series A, **164**, 517–547.

Grambsch, P.M. and T.M. Therneau (1994) Proportional hazards tests and diagnostics based on weighted residuals, *Biometrica* **81** (3), 515–526.

Greene, W.A. (2000) *Econometric Analysis* (New York: Macmillan).

Heckman, J. and B. Singer (1984) Econometric Duration Analysis, *Journal of Econometrics* **24**, 63–132.

Jenkins, S.P. (1997) Discrete time proportional hazards regression (pgmhaz). Stata Technical Bulletin STB-39, in: H.J. Newton, ed., *Stata Technical Bulletin Reprints*, Vol. 7 (College Station TX: Stata Corporation) 109–121.

Jenkins, S.P. (2004) Survival analysis. *Draft lecture notes & Introduction to the analysis of spell duration data* (Institute for Social and Economic Research, University of Essex) http://www.iser.essex.ac.uk/teaching/stephenj/ec968/index.php.

Jones, A.M. (1989) A double-hurdle model of cigarette consumption, *Journal of Applied Econometrics* **4**, 23–39.

Kiefer, N.M. (1988) Economic Duration Data and Hazard Functions, *Journal of Economic Literature* **26**, 646–679.

Klein, J.P. and M.L. Moeschberger (1997) *Survival Analysis* (New York: Springer-Verlag).

Labeaga, J.M. (1999) A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco, *Journal of Econometrics* **93**, 49–72.

Lancaster T. (1990) *The Econometric Analysis of Transition Data* (Cambridge University Press).

Lin, D.Y. and L.J. Wei (1989) The Robust Inference for the Cox Proportional Hazards Model, *Journal of the American Statistical Association* **84**, 1074–1078.

Sander, W. (1995) Schooling and quitting smoking, *Review of Economics and Statistics* **129**, 191–198.

Suranovic, S.M., R.S. Goldfarb, and T.C. Leonard (1999) An economic theory of cigarette addiction, *Journal of Health Economics* **18** (1), 1–29.

Tauras, J.A. (1999) The Transition to Smoking Cessation: Evidence from Multiple Failure Duration Analysis, *NBER Working Paper*, No. 7412.

Tauras, J.A., P.M. O'Malley, and L.D. Johnston (2001) Effects of Price and Access Laws on Teenage Smoking Initiation: A National Longitudinal Analysis, *NBER Working Paper*, No. 8331.